

RECEIVED
OCT 04 1996
OSTI

RECEIVED
OCT 01 1996
OSTI

enresa
empresa nacional de residuos radiactivos, s.a.

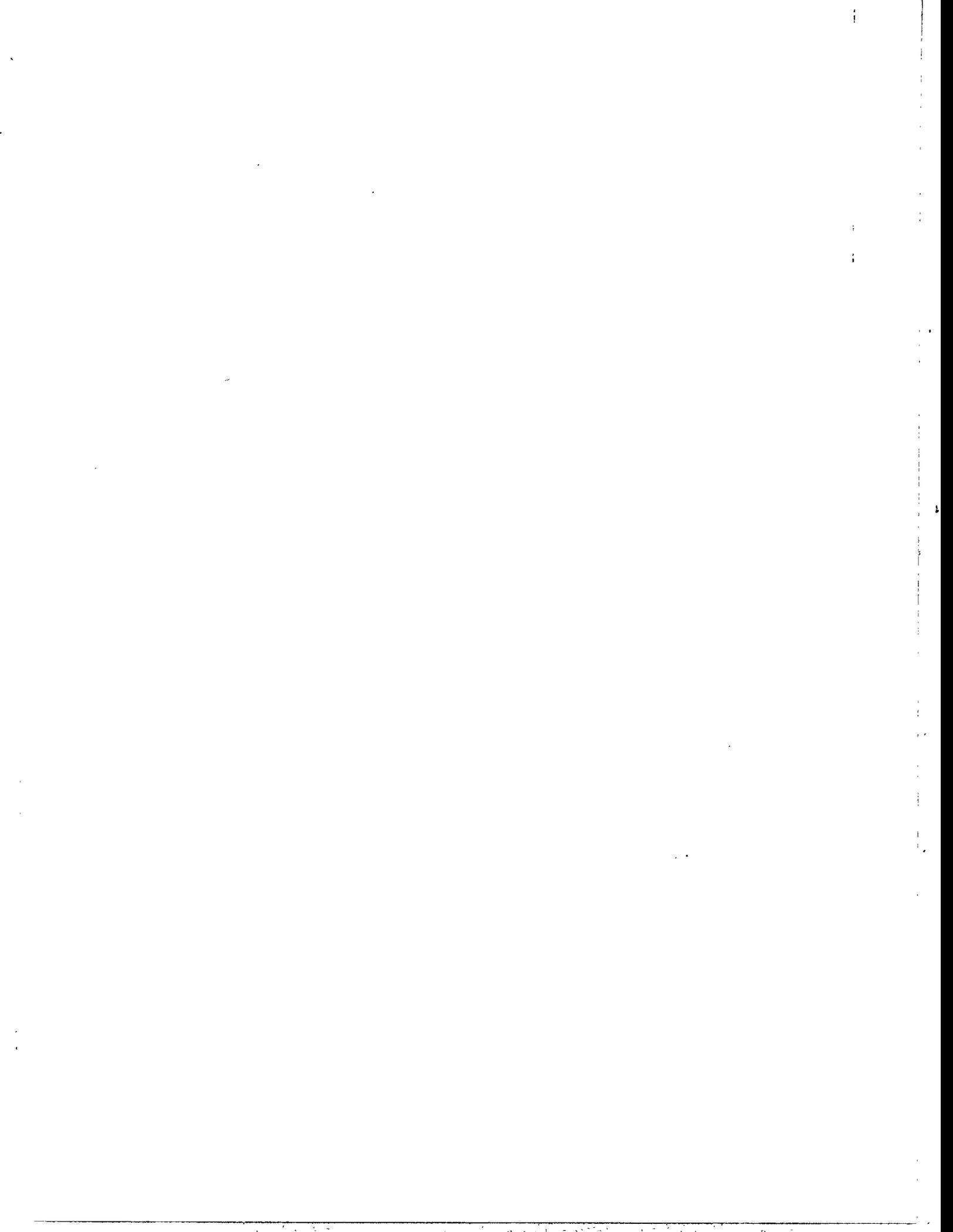
ENRESA--04/96

**METODOS GEOESTADISTICOS
PARA LA INTEGRACION
DE INFORMACION**

MASTER

PUBLICACION TECNICA NUM. 04/96

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED



METODOS GEOESTADISTICOS PARA LA INTEGRACION DE INFORMACION

Eduardo F. Cassiraga y J. Jaime Gómez-Hernández

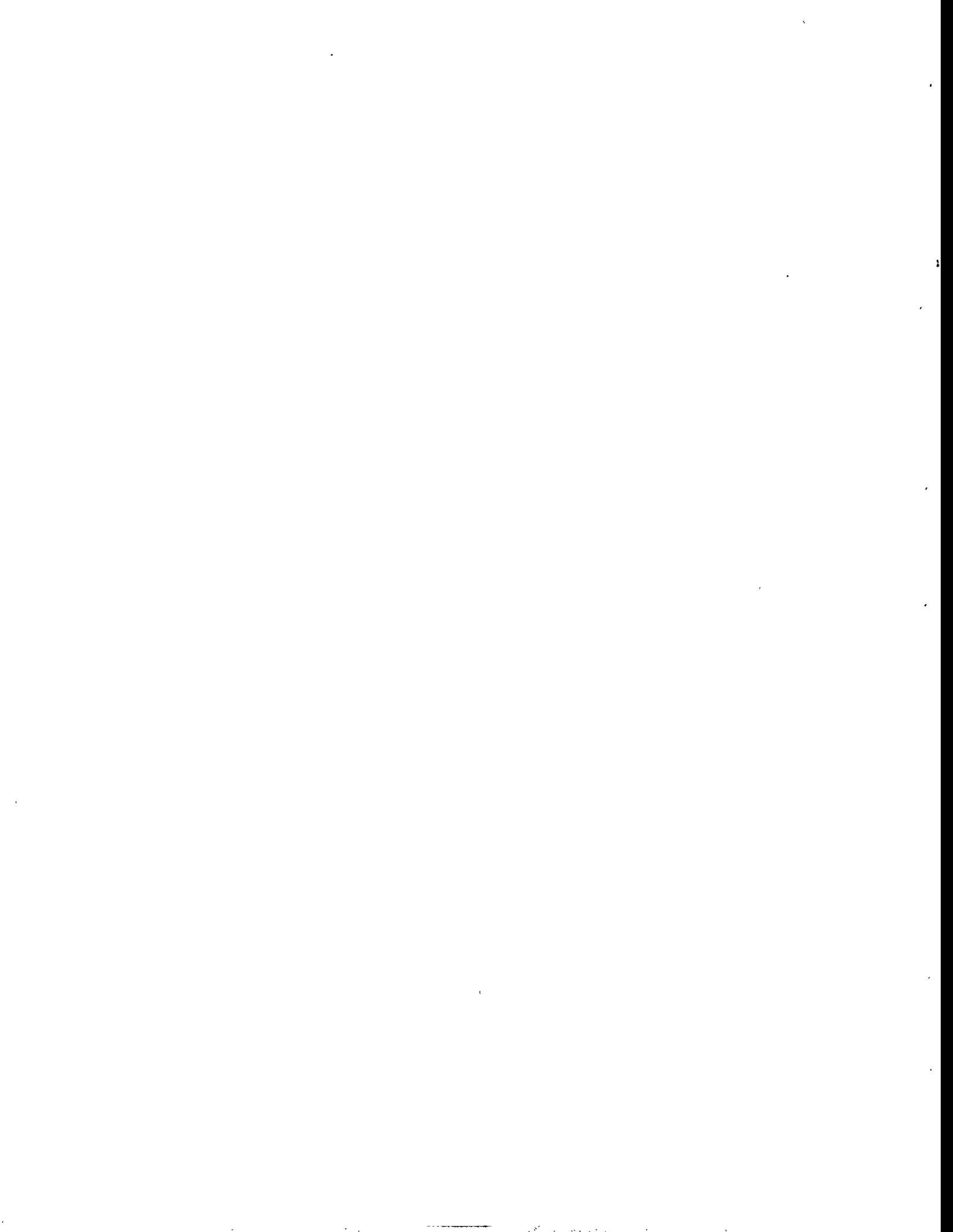
*Departamento de Ingeniería Hidráulica y Medio Ambiente
Universidad Politécnica de Valencia*

enresa



DISCLAIMER

**Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.**

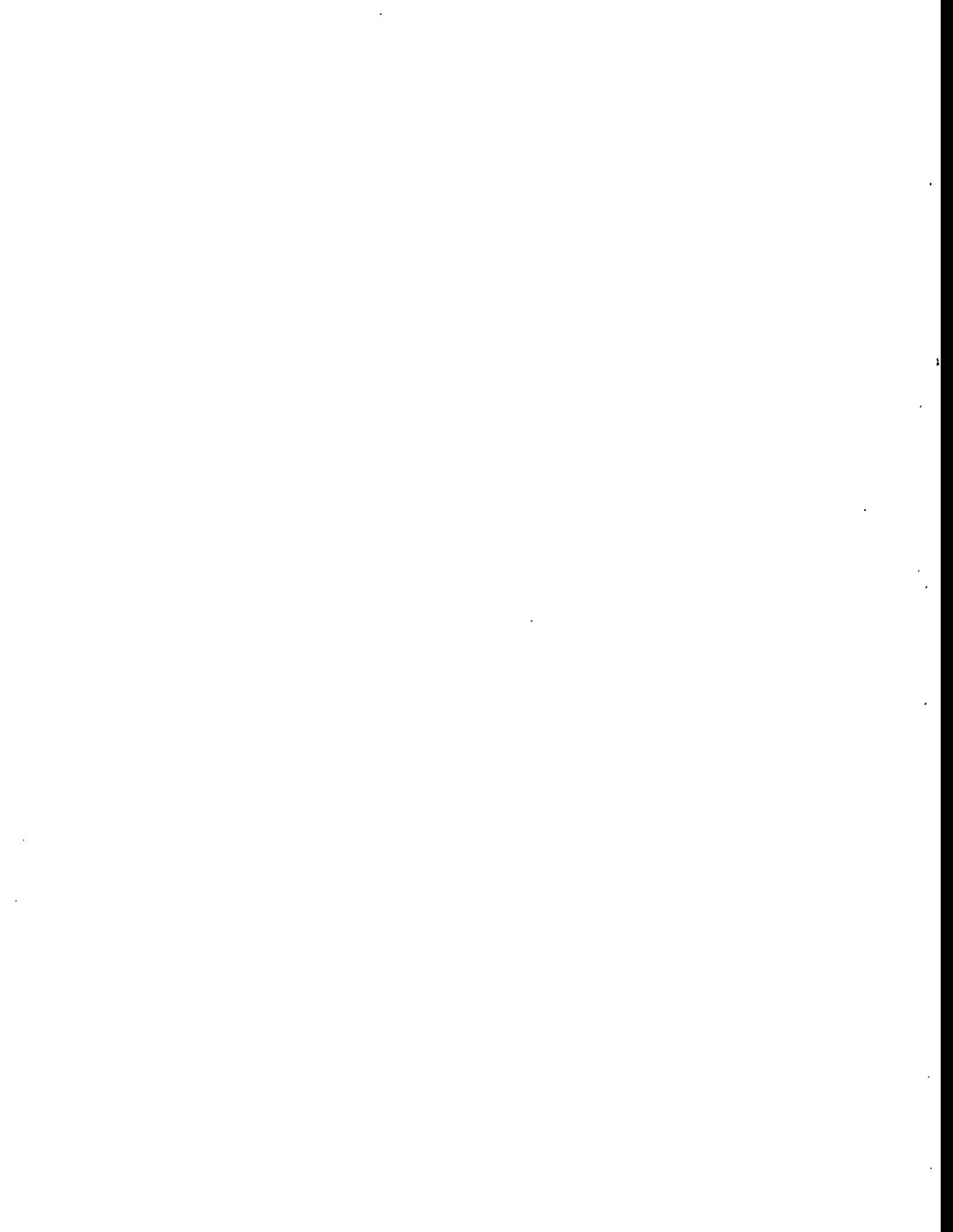


METODOS GEOESTADISTICOS PARA LA INTEGRACION DE INFORMACION

*Este trabajo ha sido realizado bajo contrato con ENRESA.
Las conclusiones y puntos de vista expresados en él corresponden a sus autores
y pueden no coincidir necesariamente con los de ENRESA.*



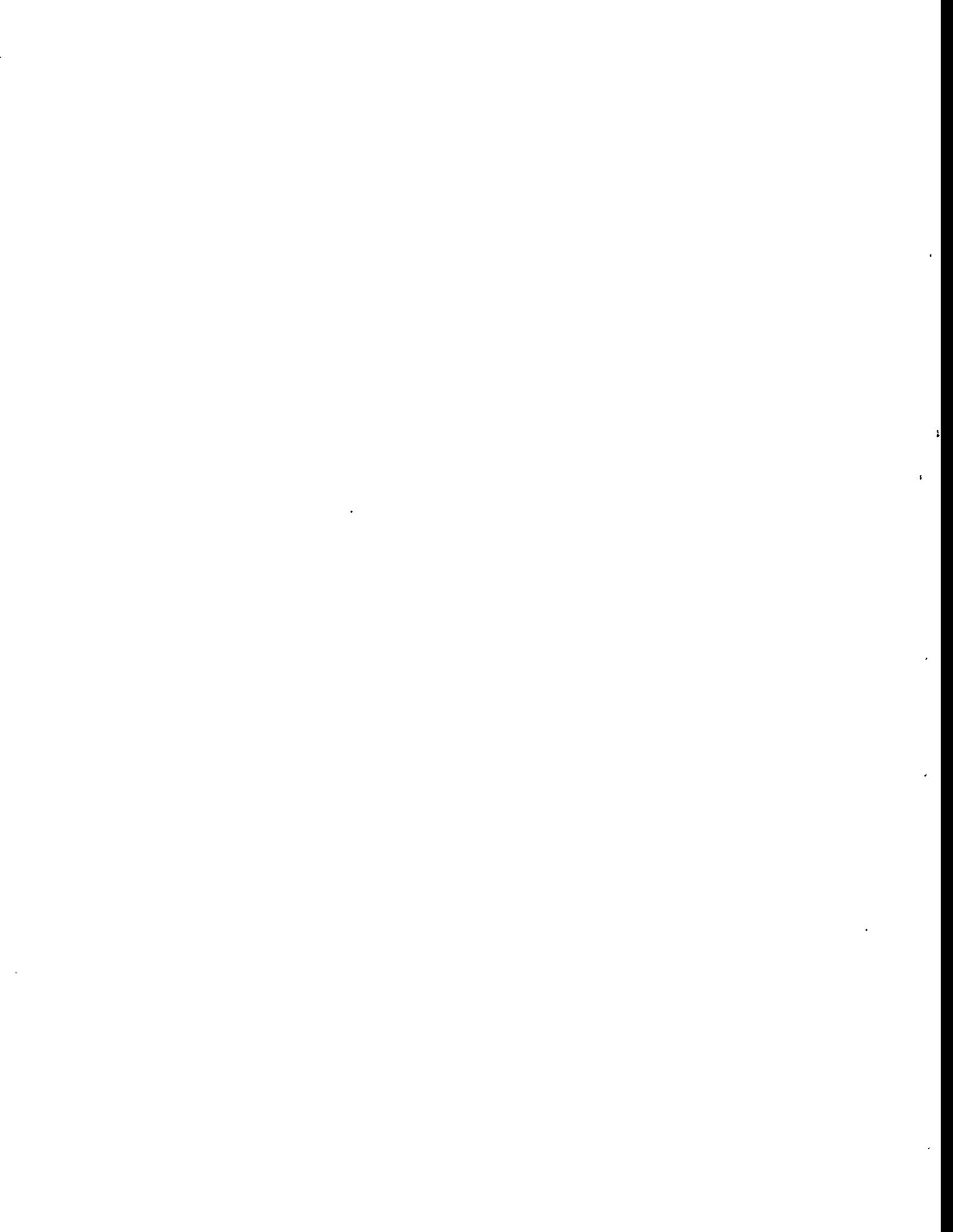
INDICE



PREAMBULO.....	1
1. INTRODUCCION	5
2. TIPOS DE INFORMACION DISPONIBLE	9
3. INTERPOLACION EN EL ESPACIO	13
4. CLASIFICACION DE LOS ALGORITMOS	17
5. ALGORITMOS DE INTERPOLACION	21
5.1 <i>Regresión tradicional</i>	23
5.2 <i>Krigeado simple, ordinario y universal</i>	23
5.3 <i>Krigeado con deriva externa</i>	24
5.4 <i>Cokrigeado simple y ordinario</i>	25
5.5 <i>Cokrigeado colocalizado</i>	27
5.6 <i>Cokrigeado colocalizado bajo un modelo de Markov</i>	27
5.7 <i>Krigeado indicador</i>	28
5.8 <i>Cokrigeado indicador bajo un modelo de Markov-Bayes</i>	30
6. ALGORITMOS DE REPRESENTACION ESTOCASTICA	35
6.1 <i>Simulación secuencial</i>	37
6.1.1 <i>Simulación gaussiana secuencial</i>	37
6.1.1.1 <i>Simulación gaussiana secuencial por cokrigeado completo</i>	38
6.1.1.2 <i>Simulación gaussiana secuencial por cokrigeado colocalizado bajo un modelo markoviano</i>	38
6.1.1.3 <i>Simulación conjunta de varias variables</i>	38
6.1.2 <i>Simulación indicadora secuencial</i>	39
6.2 <i>Simulación por recocido simulado (annealing)</i>	39
6.3 <i>Simulación por campo de probabilidades</i>	41
6.4 <i>Algoritmos booleanos</i>	42

7. EJEMPLO DE APLICACION	43
7.1 <i>Resultados de los algoritmos de estimación</i>	45
7.1.1 <i>Análisis de los resultados</i>	46
7.2 <i>Resultados de los algoritmos de simulación</i>	47
7.2.1 <i>Análisis de los resultados</i>	48
8. CONCLUSIONES	69
9. ADENDUM	73
10. REFERENCIAS	77

PREAMBULO



Este es el segundo de los informes publicados del proyecto de colaboración entre ENRESA y la Universidad Politécnica de Valencia (UPV) "Incorporación de medidas geofísicas en la estimación de permeabilidades"

El objetivo de este informe es describir una serie de técnicas geoestadísticas que nos permitan utilizar las relaciones entre parámetros geofísicos e hidrogeológicos descritas en el informe anterior,

para la modelización de la variabilidad espacial de estos últimos con el mayor grado de confianza posible.

Este informe describe con mucho mayor detalle, las características de los métodos de estimación que se utilizaron en uno de los pósteres y en la ponencia presentadas por el Departamento de Ingeniería Hidráulica y Medio Ambiente de la UPV en las segundas jornadas de I+D.



1

INTRODUCCION



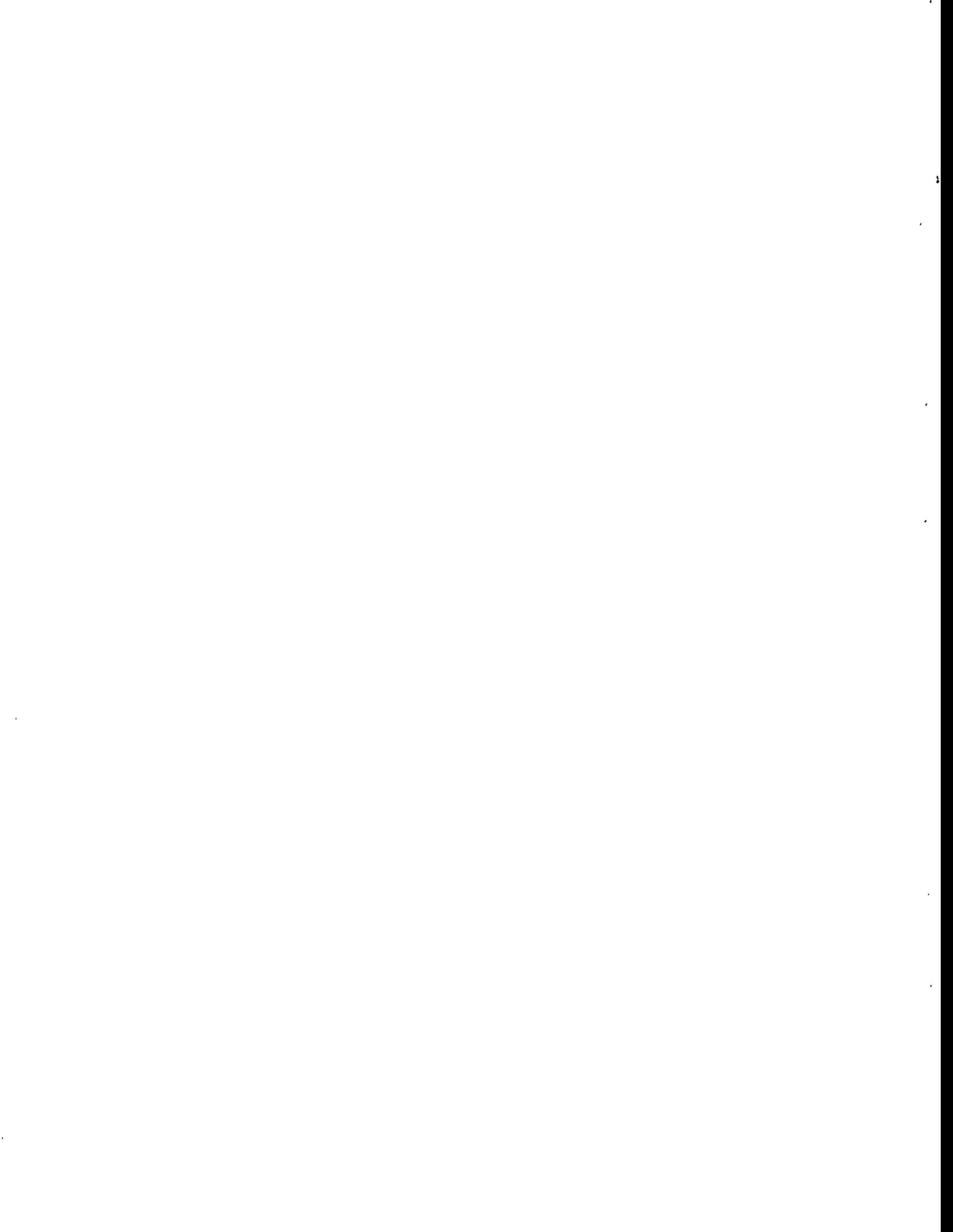
Una cuestión importante en hidrogeología, minería, industria del petróleo y medio ambiente es la consideración de la información disponible a partir de distintas fuentes, con vistas a realizar predicciones de aquellos parámetros que caracterizan el comportamiento del medio geológico estudiado.

La resolución de los problemas más comunes en cualquiera de las áreas mencionadas, requiere una buena descripción del medio en términos de arquitectura geológica, distribución espacial de heterogeneidades, zonas preferenciales de flujo y propiedades e interacciones entre rocas y fluidos. La solución pasa comúnmente por desarrollar un modelo numérico que permita predecir las respuestas del sistema, sea este un acuífero, un repositorio de residuos tóxicos, un yacimiento o una mina. Todos estos modelos se apoyan en la información proveniente de observaciones reales, y su aplicabilidad está restringida por la escasez de datos.

La geoestadística ha abordado en los últimos años la cuestión general de la integración de información, proponiendo una serie de técnicas que permiten paliar la escasez de datos mediante la incorporación de información imprecisa y, en algu-

nos casos, relacionada sólo débilmente con el parámetro en estudio. La descripción de estas técnicas es el objeto de este informe. No pretendemos presentar un desarrollo matemático exhaustivo de cada una de ellas, sino más bien describir y señalar sus ventajas e inconvenientes, así como hablar de algunos aspectos relacionados con la aplicación de las mismas a la estimación de parámetros hidrogeológicos como la porosidad y la permeabilidad, y de la geometría de la formación.

Comenzamos describiendo los distintos tipos de información con los que se puede contar en una situación real, sus fuentes y las escalas de las que provienen. Luego introducimos el problema de la interpolación de un atributo correlacionado en el espacio, con el objetivo de obtener la máxima información a partir de los datos disponibles. Sugerimos una clasificación de los algoritmos propuestos y pasamos a describir cada uno de ellos. Por último, presentamos un ejemplo de aplicación para evaluar el comportamiento de cada algoritmo. Todos los programas utilizados en este ejercicio pertenecen a la librería GSLIB (Deutsch y Journel, 1992a).



2

TIPOS DE INFORMACION DISPONIBLES



Supongamos que estamos interesados en cartografiar la porosidad de un determinado medio geológico. Inicialmente es posible contar con una serie de medidas directas de porosidad. Estas medidas pueden provenir de ensayos realizados in situ o sobre testigos en laboratorio. En ambos casos, su determinación requiere la realización de perforaciones en el terreno, lo cual implica, por un lado, un alto costo, y por otro la alteración del subsuelo. Por esta razón las medidas directas de parámetros hidrogeológicos suelen ser escasas y espacialmente dispersas. Desde el punto de vista de la modelización matemática de nuestro problema, se considera a la información directa como información precisa o dura, es decir como información cuya incertidumbre puede considerarse despreciable.

Afortunadamente, además de la información directa de porosidad, es común contar con otro tipo de información, como por ejemplo la contenida en un mapa de velocidades o atenuaciones sísmicas proveniente de la inversión de un tomograma entre pozos. Esta información se dice indirecta pues no proporciona una medida directa de la porosidad, sino la de un atributo como la velocidad o la atenuación sísmica, que esperamos tenga alguna correlación con aquélla. Dada la naturaleza indirecta de esta información se la denomina información imprecisa o blanda. Hablamos de una información cuyos valores deben ser acompañados de una medida que especifique su grado de confianza.

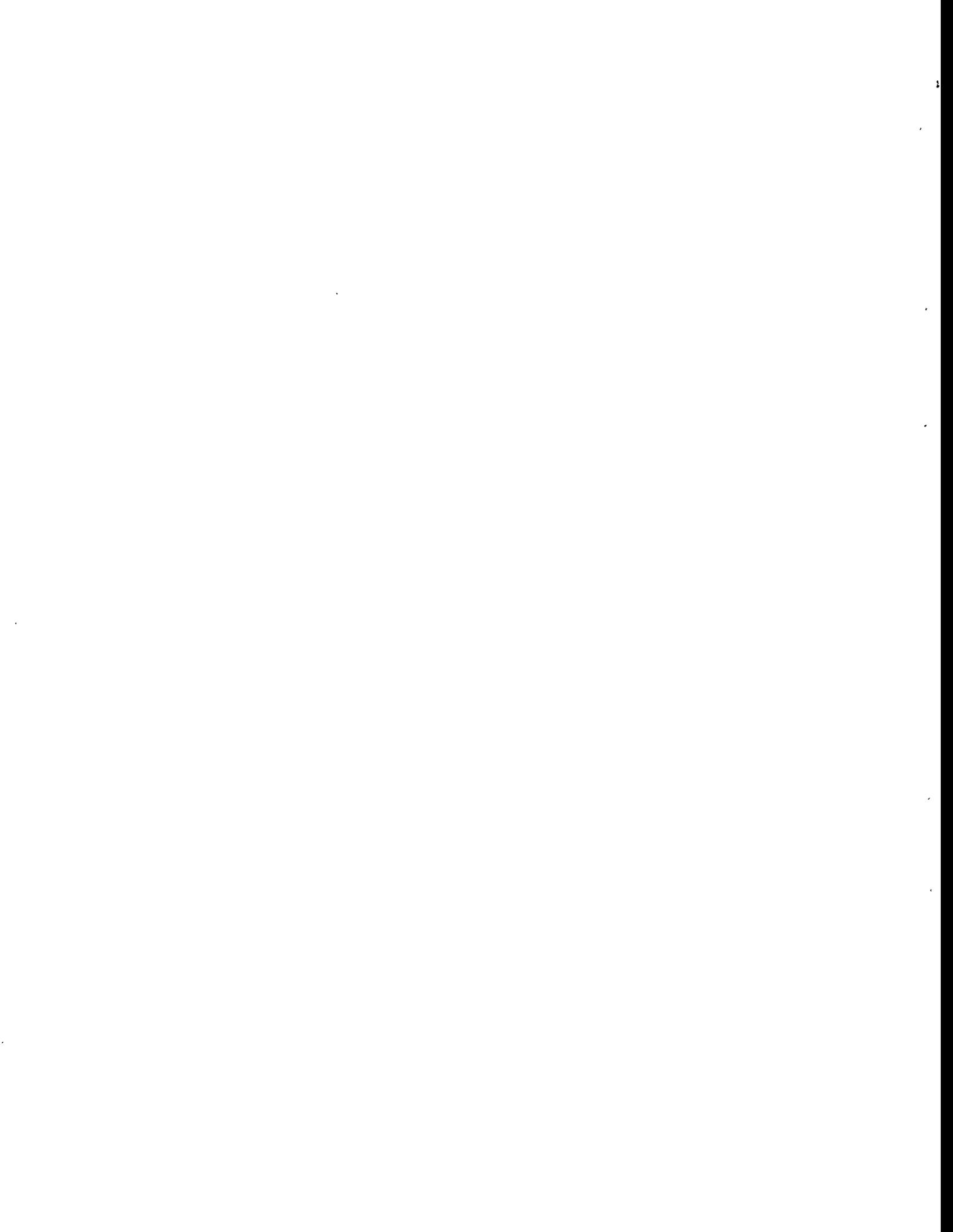
Cuando se habla de información precisa o imprecisa no nos estamos refiriendo a la bondad con la que fue medida, sino que estamos distinguiendo entre dos conjuntos de datos, uno de ellos con medidas directas del atributo que nos interesa, el otro con medidas de algún parámetro el cual debe ser tratado adecuadamente para ser incorporado en el proceso de estimación del primero. El error en la determinación de una medida de velocidad sísmica puede ser del 1%, mientras que la permeabilidad deducida a partir de un ensayo de bombeo puede tener un error del 15 o del 20%. Sin embargo la incertidumbre que existe acerca de la relación entre permeabilidad y velocidad sísmica, a la hora de estimar la primera en función de la segunda, hace que la confianza asignada a la velocidad sea menor que la otorgada a una medida directa de permeabilidad.

La importancia de considerar la información indirecta radica en que ésta puede estar disponible en mayor cantidad que la información hidrogeológica directa, y en que no es necesario destruir el medio para obtenerla. Una campaña sísmica en 3D, por ejemplo, puede suministrar información exhaustiva de la variable secundaria sobre toda el área estudiada.

Una característica adicional de la información disponible es la diferencia de las escalas a la que fue obtenida. Las medidas directas sobre testigos, las más caras y escasas, están definidas sobre una escala muy pequeña, del orden de unos pocos centímetros cúbicos. La geofísica de pozos proporciona medidas cuya representatividad está restringida a las regiones inmediatamente adyacentes a las paredes de la perforación. La interpretación geológica de los datos de pozos puede dar información acerca de la continuidad lateral de los estratos encontrados a lo largo de la perforación. La tomografía y los métodos sísmicos tradicionales proporcionan información sobre la región comprendida entre pozos. Los ensayos de bombeo son capaces de dar valores de la permeabilidad y del almacenamiento para volúmenes muy grandes.

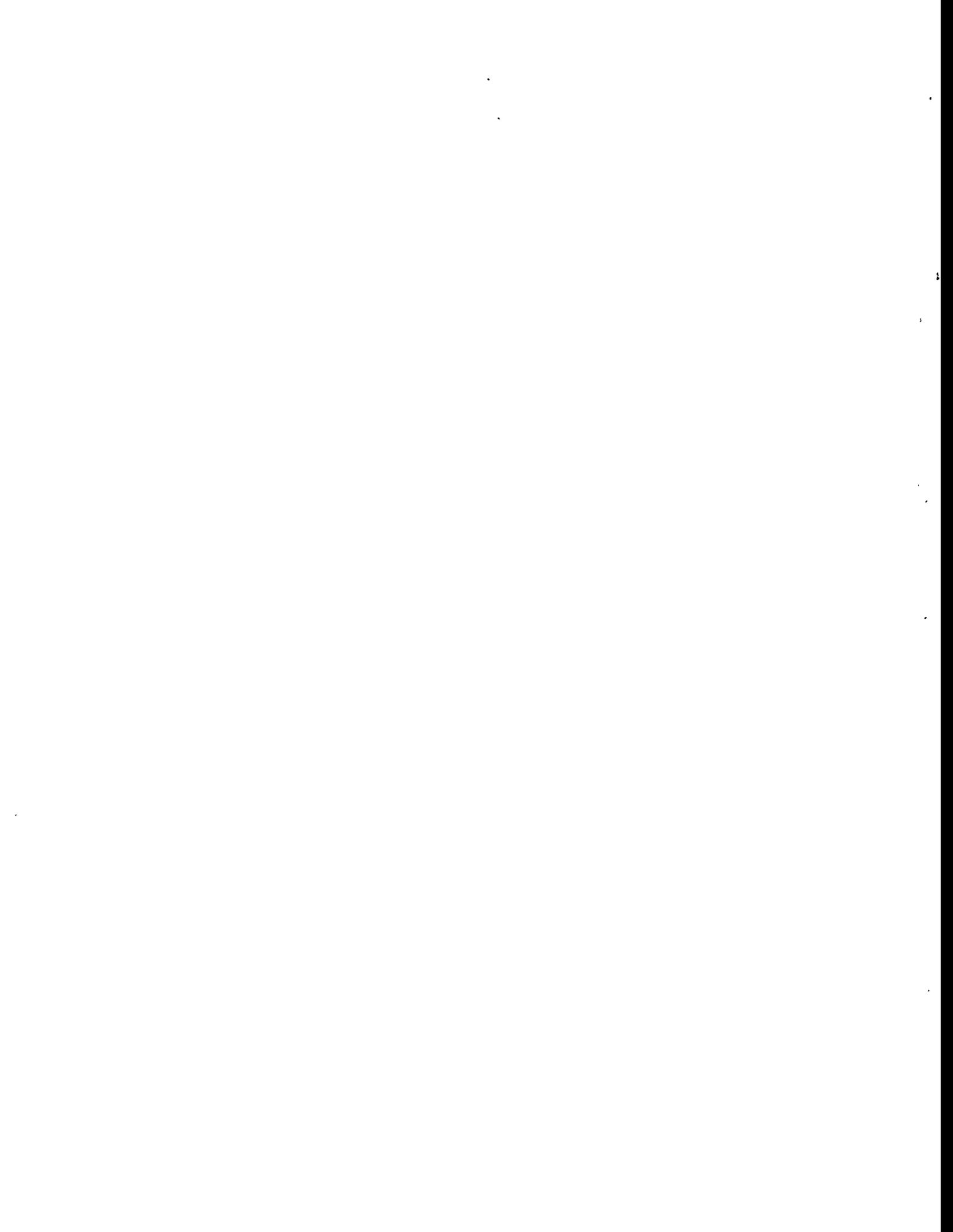
En el ejemplo que hemos introducido del cartografiado de la porosidad, los datos sísmicos son representativos de un área que puede ser de muchos metros cuadrados, mientras que la correspondiente a medidas en perforaciones es del orden de un metro cuadrado o menos. Los registros sísmicos cubren densamente la región entre pozos, mientras que los registros en perforaciones sólo están disponibles en cada una de éstas y son representativos de la región inmediatamente vecina a la pared de los pozos. Por consiguiente, una estimación de la porosidad entre pozos considerando sólo la información obtenida a partir de los registros realizados en perforaciones no es en general representativa de la realidad. Dado que la estructura espacial de la velocidad sísmica está bastante bien relacionada con la de la porosidad, la incorporación de esta información en el proceso de estimación producirá mejores resultados.

El problema es, por tanto, no sólo la integración de información de distinta calidad sino a distintas escalas, con el fin de obtener una mejor caracterización de la distribución espacial de los parámetros hidrogeológicos fundamentales.



3

INTERPOLACION EN EL ESPACIO



Se plantea el problema de la interpolación espacial de una variable a partir de datos de diferente tipo y fiabilidad, provenientes de fuentes y escalas distintas. La idea clave es utilizar todos los datos disponibles para elaborar un modelo de correlación espacial entre las variables de interés. Este modelo es exportado a zonas donde es necesario inferir la variable principal a partir de los, generalmente escasos, datos disponibles de dicha variable y de la más numerosa, pero menos precisa, información acerca de una o más variables secundarias. La eficiencia de este procedimiento depende de cuánta información es extraída de los datos, y de la representatividad del modelo de correlación espacial a exportar.

En el proceso de integración de información la calibración de los distintos tipos de datos es un paso esencial. Para realizar esta calibración es necesario contar con un conjunto de datos integrado por medidas de la variable de interés, y de la o las variables secundarias en las mismas localizaciones o muy cercanas a aquellas en donde se conoce la variable principal. Todos los parámetros que definen el modelo de correlación espacial al que hacemos referencia en el párrafo anterior son extraídos de este grupo de datos (coeficiente de correlación, covarianzas, variogramas, etc.).

En general (ver Cassiraga y Gómez-Hernández, 1994), hasta no hace mucho tiempo sólo se extraían los estadísticos de regresión lineal a partir de los datos de calibración. Sin embargo hay casos en donde una mera regresión lineal es insuficiente para cuantificar la relación entre dos o más variables.

Por ejemplo observemos el diagrama de dispersión de la figura 1 en el que se representan gráficamente los valores correspondientes a 960 parejas de calibración. A partir de este diagrama puede verse que no sólo la regresión entre las variables no es lineal, sino que la dispersión vertical de los valores es diferente según el valor considerado de la variable secundaria. Los valores bajos de ambas variables están mejor correlacionados que los valores altos.

Un modelo que será exportado a zonas no muestreadas y del cual se derivarán estimaciones de Z a partir de V , debe ser capaz de informar acerca de la dispersión señalada arriba. Es decir, no sólo una medida de la localización media de la nube de puntos es necesaria, sino también una medida de su dispersión, esto es, una varianza.

Cuando las distribuciones de Z condicionadas a los valores de V son muy diferentes, podemos estar interesados en exportar no sólo la media y la varianza de cada distribución, sino también la forma de la misma. Nos interesa entonces modelizar la distribución completa de la variable a estimar condicionada a la información secundaria, la cual se expresa como:

$$\text{Prob} \{Z \leq z_k | V \in (v_1, v_2)\}$$

donde Z representa la variable principal, z_k un valor dado de ésta, V la variable secundaria, y (v_1, v_2) un intervalo en el rango de variación de V .

La expresión anterior representa la información que se tiene acerca de la variable principal cuando la secundaria pertenece a un intervalo dado. Las probabilidades asociadas a la expresión anterior pueden derivarse a partir de un conjunto de datos de calibración.

En un esquema de análisis bayesiano (ver Box y Tiao, 1973) estos valores definen una función de probabilidad previa calibrada a partir de un mínimo de información (los datos de calibración). El objetivo del problema de interpolación en el espacio es mejorar esta función inicial incorporando las medidas disponibles de las variables consideradas, tanto duras como blandas, en cada localización y en su entorno. Esto da como resultado una función de probabilidad posterior, función que está condicionada a toda la información disponible. Esto es:

$$\text{Prob} \{Z(u) \leq z_k | \text{ toda la información disponible}\}$$

Una función como la descrita arriba constituye un modelo de incertidumbre de la variable de interés. Si disponemos de modelos de incertidumbre en cada localización a estimar es posible obtener mapas de riesgo representando distintos índices de interés, por ejemplo, un estimador óptimo de la variable, probabilidades de que su valor esté por encima o por debajo de un cierto umbral, o el error cometido en las estimaciones. (Para una descripción del concepto de modelos de incertidumbre y su utilización en la confección de mapas de riesgo ver Gómez-Hernández 1991, 1993a y b).

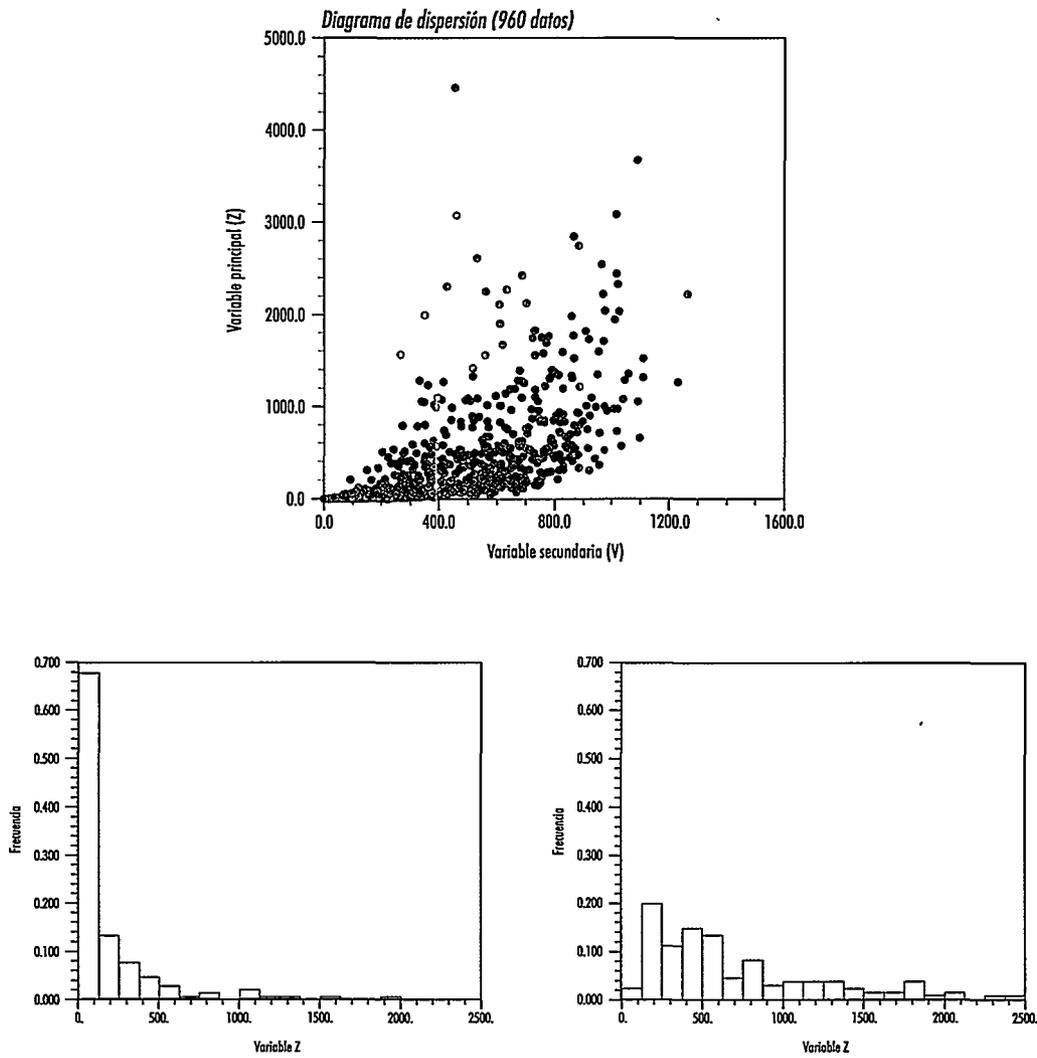


Figura 1. Ejemplo de un diagrama de dispersión en el cual la regresión es no lineal. Las distribuciones condicionales de Z cambian de forma, de sesgo y de varianza según las clases de V. Los histogramas corresponden a los puntos de dentro de las barras sombreadas para $V \in a (200;400)$ a la izquierda y para $V \in (600;800)$ a la derecha.

4

CLASIFICACION DE ALGORITMOS



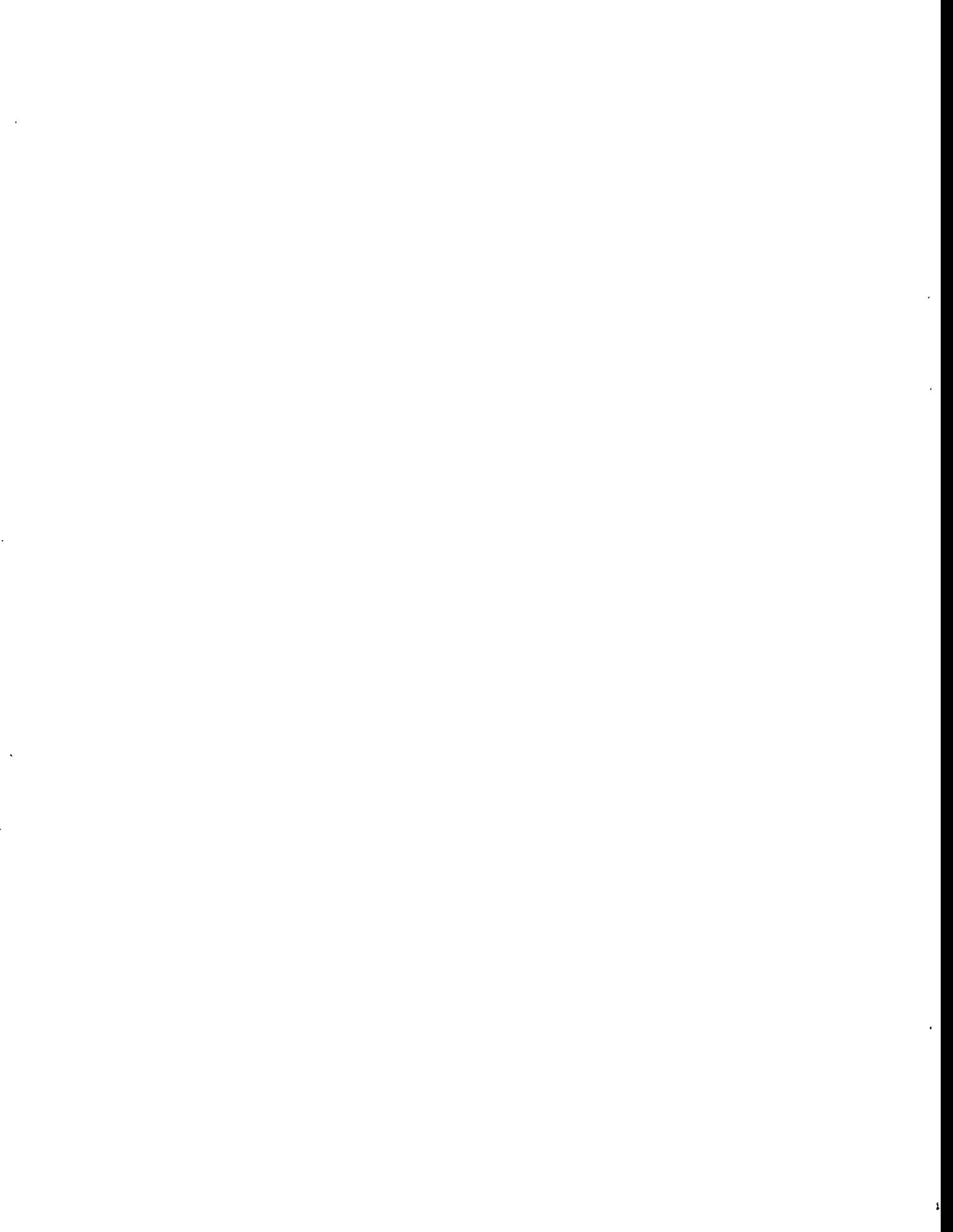
Los algoritmos para cartografiar una variable principal a partir tanto de la información dura disponible como de la información blanda acerca de una o más variables secundarias, pueden clasificarse en dos grupos: algoritmos de *interpolación o estimación*, y algoritmos de *representación estocástica o simulación* (para un tratamiento más extenso de estas dos categorías el lector puede referirse a Gómez-Hernández, 1991).

Los algoritmos de interpolación producen una única imagen que consiste en un campo interpolado de la variable de interés. En cada nodo de ese campo se estima un valor de acuerdo a un cierto criterio de optimalidad. Por ejemplo en el kriging/cokriging se busca minimizar la varianza del error cuadrático medio cometido en cada estimación. Esta varianza mínima es considerada una medida de incertidumbre de la estimación realizada, y el mapa construido con ellas es indicativo de la incertidumbre del campo estimado. Los campos

obtenidos por cualquier algoritmo de interpolación suelen ser suaves, no reflejando la variabilidad que el parámetro estimado tiene en la realidad. Cada valor estimado es "coherente" con los datos pero no con los valores estimados en otros nodos.

Las técnicas de representación estocástica persiguen la generación de realizaciones múltiples del atributo simulado, todas igualmente plausibles en el sentido de que reproducen los datos y los patrones de correlación espacial observados.

La información dura se reproduce exactamente y la blanda con un grado de tolerancia determinado por la precisión de la misma. Al generar varias realizaciones del mismo proceso, las diferencias entre cada realización proveen una medida cuantitativa y visual de la incertidumbre en el parámetro modelizado. Puesto que todas las realizaciones reproducen los datos y el grado de variabilidad impuesto, cualquier realización podría corresponder a la realidad.



5

ALGORITMOS DE INTERPOLACION



5.1 Regresión tradicional

La predicción más simple de una variable en función de la otra es la regresión lineal. En tal modelo asumimos que la dependencia entre variables puede ser adecuadamente descrita por una línea recta, es decir, por una ecuación del tipo:

$$z_1 = a z_2 + b$$

Una técnica para estimar los parámetros a y b en este modelo, es el método de los mínimos cuadrados. Después de aplicar este método la pendiente a y la constante b resultan tener las expresiones siguientes:

$$a = \rho \frac{\sigma_{z_2}}{\sigma_{z_1}} \quad b = m_{z_2} - a m_{z_1}$$

donde ρ es el coeficiente de correlación lineal entre las variables z_1 y z_2 ; σ_{z_1} es la desviación típica de los valores de la variable z_1 ; σ_{z_2} es la de los valores de la variable z_2 ; y m_{z_1} y m_{z_2} sus correspondientes valores medios.

Este algoritmo es extremadamente sencillo aunque presenta limitaciones importantes que se describen a continuación.

En cada punto donde se quiere estimar la variable principal, debemos contar con una medida de la variable secundaria. Los valores estimados son función de un sólo valor, cuando es más razonable pensar que un parámetro correlacionado en el espacio, lo sea también de los valores en su entorno. En esta técnica no existe la posibilidad ni de cuantificar la incertidumbre de las estimaciones realizadas, ni de tener en cuenta la correlación espacial de los datos.

Por otro lado la recta ajustada no reproduce, en general, los datos disponibles. Al calibrar un grupo de datos según un modelo de regresión lineal como el descrito, sólo estamos utilizando la información contenida en la media. La técnica de regresión lineal por mínimos cuadrados ha sido ampliamente utilizada por algunos investigadores para la integración/correlación de parámetros geoelectrónicos e hidrogeológicos (ver la revisión realizada por Cassiraga y Gómez-Hernández, 1994), y también fue aplicada con datos sísmicos (Doyen, 1988).

5.2 Krigado simple, ordinario y universal

Aunque el krigado (ver Journel, 1989; Isaaks y Srivastava, 1989; y Deutsch, 1991), permite sólo la consideración de una variable, lo describiremos aquí ya que está en la base de las técnicas que introduciremos más adelante.

Sea $z(u)$ el parámetro de interés, donde u es el vector que indica sus coordenadas espaciales, para el que adoptamos un modelo estocástico. Este modelo estocástico consiste en admitir que los valores que puede tomar z dentro del área en estudio vienen dados por una función aleatoria $Z(u)$. La idea básica del krigado es estimar el valor desconocido de z en el punto de coordenadas u_0 como una combinación lineal de n valores conocidos de $z(u_\alpha)$, $\alpha=1, \dots, n$. De acuerdo a esto, el estimador por krigado simple es:

$$z_{KS}^*(u_0) - m = \sum_{\alpha=1}^n \lambda_\alpha [z(u_\alpha) - m]$$

donde el símbolo * denota que se trata de un estimador, m es igual al valor esperado de la variable aleatoria Z , que en la práctica se identifica con el valor medio de los datos; y λ_α , $\alpha=1, \dots, n$ son coeficientes de ponderación. Estos coeficientes vienen determinados al exigir que el estimador sea insesgado y que la varianza de los errores sea mínima, lo cual conduce al siguiente sistema lineal de ecuaciones, llamado sistema de krigado simple:

$$\sum_{\beta=1}^n \lambda_\beta C(u_\alpha - u_\beta) = C(u_\alpha - u_0), \quad \alpha = 1, \dots, n$$

donde $C(h)$ es la covarianza para el vector de separación h . A la izquierda de la igualdad están las covarianzas entre los datos y a la derecha entre los datos y el punto a estimar. La varianza de los errores por krigado simple resulta:

$$\sigma_{KS}^2(u_0) = \sigma^2 - \sum_{\alpha=1}^n \lambda_\alpha C(u_\alpha - u_0)$$

donde σ^2 es la varianza correspondiente a los datos que intervienen en la estimación. El mapa de los valores de σ_{KS}^2 permite cuantificar la incertidumbre en el campo estimado. El estimador por krigado simple requiere el conocimiento previo de la media m . En la formulación presentada se

ha supuesto que el valor m es constante sobre todo el campo estimado.

Otra posibilidad dentro del krigado simple surge si contamos con un grupo de datos de calibración, a partir del cual sea posible estimar un valor medio local de la variable principal. Esta media se lee directamente del diagrama de dispersión de los datos de calibración, para lo cual se discretiza el rango de variación de la variable secundaria y se calcula la media de los valores de la variable principal dentro de cada intervalo de aquella. Hablamos entonces de *krigado simple con media variable*.

En el *krigado ordinario*, la media es desconocida y se reestima localmente a partir de los datos que intervienen en cada estimación. El *estimador por krigado ordinario* esta dado por la siguiente expresión:

$$z_{KO}^*(u_0) = \sum_{\alpha=1}^n \lambda_{\alpha} z(u_{\alpha})$$

La sumatoria de los pesos debe ser igual a la unidad para que el estimador obtenido sea insesgado, que junto a la condición de error mínimo resulta en el *sistema de krigado ordinario*:

$$\sum_{\beta=1}^n \lambda_{\beta} C(u_{\alpha} - u_{\beta}) + \mu = C(u_{\alpha} - u_0), \quad \alpha = 1, \dots, n$$

$$\sum_{\beta=1}^n \lambda_{\beta} = 1$$

donde μ es el parámetro de Lagrange introducido en la minimización del error sujeto a la condición de que los pesos sumen uno. La expresión para la *varianza de los errores por krigado ordinario* es la siguiente:

$$\sigma_{KO}^2(u_0) = \sigma^2 - \left[\sum_{\alpha=1}^n \lambda_{\alpha} C(u_{\alpha} - u_0) + \mu \right]$$

El *krigado universal*, o mejor dicho el *krigado con un modelo de tendencia*, considera que la variación espacial de la media local tiene una forma dada por un polinomio de orden bajo en función de las coordenadas, pero de parámetros desconocidos. En general esta tendencia tiene la forma siguiente:

$$m(u) = \sum_{l=0}^L a_l f_l(u)$$

donde $m(u)$ es la media en el punto de coordenadas u ; a_l , con $l=0, \dots, L$ son coeficientes desconocidos, y $f_l(u)$ son monomios de bajo orden función de las coordenadas. Por ejemplo un modelo lineal en dos dimensiones sería:

$$m(x, y) = a_0 + a_1 x + a_2 y$$

El *estimador por krigado universal* es idéntico al de *krigado ordinario*, es decir:

$$z_{KU}^*(u_0) = \sum_{\alpha=1}^n \lambda_{\alpha} z(u_{\alpha})$$

La condición de no sesgo resulta en L restricciones que deben satisfacer los pesos, que unidas a la minimización del error resulta en el *sistema de krigado universal*, cuyas expresiones son las siguientes:

$$\sum_{\beta=1}^n \lambda_{\beta} C(u_{\alpha} - u_{\beta}) + \sum_{l=0}^L \mu_l f_l(u_{\alpha}) =$$

$$= C(u_{\alpha} - u_0), \quad \alpha = 1, \dots, n$$

$$\sum_{\beta=1}^n \lambda_{\beta} f_l(u_{\beta}) = f_l(u_0), \quad l = 0, \dots, L$$

La 0-ésima restricción es la clásica restricción del *krigado ordinario*, que corresponde a un modelo $m(x, y) = a_0$. La *varianza de los errores por krigado universal* viene dada por la siguiente expresión:

$$\sigma_{KU}^2(u_0) = \sigma^2 - \left[\sum_{\alpha=1}^n \lambda_{\alpha} C(u_{\alpha} - u_0) + \sum_{l=0}^L \mu_l f_l(u_{\alpha}) \right],$$

$$\alpha = 1, \dots, n$$

5.3 Krigado con deriva externa

El *krigado con una deriva externa* (ver Journel y Rossi, 1989 y Deutsch, 1991) es una extensión del *krigado universal*. Consiste en considerar una función de tendencia $f_1(u)$, definida en cada localización a partir de una variable secundaria $z_2(u)$, que debe variar suavemente en el espacio. Es ne-

cesario disponer de un valor de la variable secundaria en cada punto a estimar. Si éste no fuera el caso, una solución sería krigear primero la variable secundaria en todo el dominio del problema.

En general se considera a la variable secundaria $z_2(u)$ relacionada de forma lineal con la media de la variable primaria $m_1(u)$, tal que:

$$m_1(u) = a + b \cdot z_2(u)$$

El *estimador por krigado con deriva externa* es idéntico al de krigado ordinario y su expresión es:

$$z_{KDE}^*(u_0) = \sum_{\alpha=1}^{n_1} \lambda_{\alpha} z_1(u_{\alpha})$$

donde n_1 es el número de datos correspondientes a la variable primaria $z_1(u)$ que intervienen en cada estimación. Los coeficientes en este estimador surgen de resolver el sistema de krigado con una deriva externa cuyas ecuaciones son las siguientes:

$$\sum_{\beta=1}^{n_1} \lambda_{\beta} C_1(u_{\alpha} - u_{\beta}) + \mu_1 + \mu_2 z_2(u_{\alpha}) = C_1(u_{\alpha} - u_0),$$

$$\alpha = 1, \dots, n_1$$

$$\sum_{\beta=1}^{n_1} \lambda_{\beta} = 1$$

$$\sum_{\beta=1}^{n_1} \lambda_{\beta} z_2(u_{\beta}) = z_2(u_0)$$

Las dos últimas ecuaciones resultan de imponer la condición de no sesgo. Nótese que este sistema de ecuaciones sólo utiliza la covarianza de la variable principal. La expresión para la estimación de la *varianza de los errores por krigado con una deriva externa* es similar a la correspondiente al krigado universal y es la siguiente:

$$z_{KDE}^2(u_0) = \sigma^2 - \left[\sum_{\alpha=1}^{n_1} \lambda_{\alpha} C(u_{\alpha} - u_0) + \mu_1 + \mu_2 z_2(u_{\alpha}) \right]$$

El campo resultante de z_1 es por construcción semejante al campo de la variable secundaria z_2 . Sin embargo este algoritmo no considera la correlación cruzada entre variables. La variable secundaria sólo informa acerca de la deriva de la variable

primaria. Por ejemplo, y si bien esto puede ser difícil de documentar en la práctica, puede considerarse que los tiempos de viaje sísmicos definen las tendencias locales en la porosidad. Esta tendencia no puede ser inferida a partir de los escasos y dispersos datos directos de porosidad, por lo que resulta interesante incorporarla a través de krigado con una deriva externa. Marechal (1984) y Araktingi *et al.* (1992) utilizaron esta técnica con datos sísmicos.

5.4 Cokrigado simple y ordinario

Las técnicas de krigado presentadas hasta ahora sólo se diferencian en la manera en que incorporan la superficie media en un problema de interpolación espacial. Hemos visto que en el caso de krigado con una deriva externa es posible, bajo ciertas condiciones, incorporar información secundaria. El cokrigado (ver Myers, 1982, 1984 y 1985; y Isaaks y Srivastava, 1989), no se diferencia del krigado desde un punto de vista algorítmico. Su novedad es que permite que los datos secundarios participen directamente en la estimación de la variable principal. Es decir que la información secundaria se trata como información covariada.

Desde el punto de vista de la incorporación de atributos externos, el cokrigado hace mejor uso de la información en el sentido de que, tanto la magnitud de la variable secundaria, como sus patrones de variación espacial, son considerados en la estimación de la variable principal. La cantidad de codependencia lineal es capturada por las covarianzas cruzadas y, a diferencia del krigado con una deriva, la variable externa no necesita estar disponible en todos los puntos en los que se desea estimar la variable primaria.

Si disponemos de datos de las variables primaria y secundaria el *estimador por cokrigado simple* de la variable principal viene dado por la siguiente expresión:

$$z_{CKS}^*(u_0) - m_1 = \sum_{\alpha=1}^{n_1} \lambda_{\alpha}^{(1)} [z_1(u_{\alpha}) - m_1] + \sum_{\alpha=1}^{n_2} \lambda_{\alpha}^{(2)} [z_2(u'_{\alpha}) - m_2]$$

donde z_1 es la variable principal; z_2 es la variable secundaria; m_1 y m_2 son sus medias; n_1 es el número

mero de datos primarios y n_2 es el número de datos secundarios. Los pesos $\lambda_\alpha^{(1)}$ y $\lambda_\alpha^{(2)}$ se obtienen resolviendo el siguiente sistema de ecuaciones lineales o *sistema de cokrigado simple* que resulta de la minimización del error cuadrático medio:

$$\sum_{\beta=1}^{n_1} \lambda_\beta^{(1)} C_1(u_\alpha - u_\beta) + \sum_{\beta=1}^{n_2} \lambda_\beta^{(2)} C_{12}(u_\alpha - u'_\beta) =$$

$$= C_1(u_\alpha - u_0), \quad \alpha = 1, \dots, n_1$$

$$\sum_{\beta=1}^{n_1} \lambda_\beta^{(1)} C_{12}(u'_\alpha - u'_\beta) + \sum_{\beta=1}^{n_2} \lambda_\beta^{(2)} C_2(u'_\alpha - u'_\beta) =$$

$$= C_{12}(u'_\alpha - u_0), \quad \alpha = 1, \dots, n_2$$

donde intervienen tanto la covarianza de la variable primaria $C_1(u_\alpha - u_\beta)$ como la de la secundaria $C_2(u'_\alpha - u'_\beta)$ y la covarianza cruzada $C_{12}(u_\alpha - u'_\beta)$. Como en el krigado simple, el algoritmo de cokrigado simple requiere conocer las medias de las variables consideradas. La expresión para la varianza de los errores por cokrigado simple es la siguiente:

$$\sigma_{CKS}^2(u_0) = \sigma_1^2 - \sum_{\alpha=1}^{n_1} \lambda_\alpha^{(1)} C_1(u_\alpha - u_0) - \sum_{\alpha=1}^{n_2} \lambda_\alpha^{(2)} C_{12}(u'_\alpha - u_0)$$

El estimador por cokrigado ordinario está dado por la siguiente ecuación:

$$z_{CKO}(u_0) = \sum_{\alpha=1}^{n_1} \lambda_\alpha^{(1)} z_1(u_\alpha) + \sum_{\alpha=1}^{n_2} \lambda_\alpha^{(2)} z_2(u'_\alpha)$$

y los coeficientes de ponderación resultan de la solución de las ecuaciones siguientes o *sistema de cokrigado ordinario*:

$$\sum_{\beta=1}^{n_1} \lambda_\beta^{(1)} C_1(u_\alpha - u_\beta) + \sum_{\beta=1}^{n_2} \lambda_\beta^{(2)} C_{12}(u_\alpha - u'_\beta) + \mu_1 =$$

$$= C_1(u_\alpha - u_0), \quad \alpha = 1, \dots, n_1$$

$$\sum_{\beta=1}^{n_1} \lambda_\beta^{(1)} C_{12}(u'_\alpha - u_\beta) + \sum_{\beta=1}^{n_2} \lambda_\beta^{(2)} C_2(u'_\alpha - u'_\beta) + \mu_2 =$$

$$= C_{12}(u'_\alpha - u_0), \quad \alpha = 1, \dots, n_2$$

$$\sum_{\beta=1}^{n_1} \lambda_\beta^{(1)} = 1,$$

$$\sum_{\beta=1}^{n_2} \lambda_\beta^{(2)} = 0$$

donde las dos últimas restricciones son suficientes, aunque no necesarias, para garantizar una estimación insesgada, y μ_1 y μ_2 son los parámetros de Lagrange introducidos en la minimización. La expresión para la *varianza de los errores por cokrigado ordinario* es la siguiente:

$$\sigma_{CKO}^2(u_0) = \sigma_1^2 + \mu_1 - \sum_{\alpha=1}^{n_1} \lambda_\alpha^{(1)} C_1(u_\alpha - u_0) - \sum_{\alpha=1}^{n_2} \lambda_\alpha^{(2)} C_{12}(u'_\alpha - u_0)$$

En el campo de la hidrogeología el cokrigado fue aplicado por Aboufirassi y Mariño (1984) y por Muñoz-Pardo y García (1989) para estimar campos de transmisividad incorporando datos de capacidad específica. Bardossy *et al.* (1986) lo utilizaron para predecir la capacidad específica de pozos en un acuífero contando con medidas superficiales de resistividad y con un número escaso de datos registrados en perforaciones. Ahmed *et al.* (1988) lo aplicaron al cartografiado de la transmisividad a partir del conocimiento de unos pocos datos directos de ésta y la capacidad específica y de numerosos datos de resistividad eléctrica. Doyen (1988) estimó campos de porosidad a partir de datos sísmicos. Mavko *et al.* (1991) estimaron la porosidad entre dos pozos cokrigando datos de porosidad y datos de velocidades sísmicas. Pesti *et al.* (1993) delinearon la geometría de un estrato arcilloso que actúa como horizonte protector natural contra la contaminación del acuífero que lo subyace. Ahmed y De Marsily (1987) compararon algunos métodos geoesdísticos entre los que estaban el krigado con una deriva externa y el cokrigado, para la estimación de transmisividad incorporando datos de capacidad específica. Istok *et al.* (1993) lo hicieron con el krigado y cokrigado para estimar concentraciones de pesticidas en un acuífero aluvial. Asli y Marcotte (1995) con diversas variantes del krigado y el cokrigado trabajando principalmente con datos geoquímicos. Cas-

siraga y Gómez-Hernández (1995) compararon varias técnicas geoestadísticas de estimación utilizando una base de datos de génesis natural obtenidos a partir de un modelo de elevación digital.

5.5 Cokrigado colocalizado

La resolución de un sistema de cokrigado completo como el presentado en el apartado anterior, requiere el conocimiento de las funciones de covarianza de las variables principal y secundaria y la de covarianza cruzada entre ellas si una sola variable secundaria es incorporada. Esto significa más esfuerzo en términos de modelización y tiempo de ordenador respecto a las técnicas de krigado, las cuales sólo requerían la función de covarianza de la variable principal. Muchas veces, este mayor esfuerzo de cálculo, constituye la principal razón por la cual el cokrigado no es utilizado. Por otro lado, cuando la información secundaria está casi exhaustivamente muestreada, como es el caso de los datos sísmicos, existe una dificultad adicional a la hora de la implementación de un cokrigado completo, que se manifiesta a través de la inestabilidad de las matrices del sistema. Esto está causado por la extrema proximidad y fuerte autocorrelación que existe en general entre los datos blandos, y las grandes distancias y pobre correlación entre los datos duros. Estas razones, más el hecho de que si hay un dato secundario colocalizado en la localización a ser estimada, éste tiende a atenuar la influencia de los datos vecinos (efecto pantalla), nos conduce a presentar la alternativa del cokrigado colocalizado (ver Journel *et al.*, 1992 y Almeida y Journel, 1994).

Se trata de un cokrigado que como información externa retiene en cada punto de coordenadas u donde se va a estimar la variable primaria, sólo el dato secundario localizado en ese punto (dato colocalizado). El *estimador por cokrigado colocalizado simple* para la variable principal es:

$$z_1^*(u_0) - m_1 = \sum_{\alpha=1}^n \lambda_{\alpha}^{(1)} [z_1(u_{\alpha}) - m_1] + \lambda^{(2)} [z_2(u_0) - m_2]$$

y el sistema que proporciona los coeficientes de ponderación:

$$\sum_{\beta=1}^n \lambda_{\beta}^{(1)} C_1(u_{\alpha} - u_{\beta}) + \lambda^{(2)} C_{12}(u_{\alpha} - u_0) =$$

$$= C_1(u_{\alpha} - u_0), \quad \alpha = 1, \dots, n_1$$

$$\sum_{\beta=1}^n \lambda_{\beta}^{(1)} C_{12}(u_0 - u_{\beta}) + \lambda^{(2)} C_2(0) = C_{12}(0)$$

Un sistema como el anterior es idéntico al de cokrigado simple completo donde el número de datos secundarios n_2 es igual a 1. Aún así, este algoritmo requiere el cálculo de las covarianzas cruzadas $C_{12}(h)$ entre variables, con h igual a la separación entre los puntos considerados. El uso de un modelo markoviano alivia esta tarea.

5.6 Cokrigado colocalizado bajo un modelo de Markov

Este algoritmo recurre a un modelo markoviano que expresa que el dato duro $z_1(u)$ "oculta" (efecto pantalla) la influencia de cualquier otro dato duro $z_1(u+h)$ sobre el dato secundario colocalizado $z_2(u)$, es decir que:

$$E\{Z_2(u) | Z_1(u) = z, Z_1(u+h) = z'\} =$$

$$E\{Z_2(u) | Z_1(u) = z\}, \quad \forall h, z$$

A partir de esta asunción se puede demostrar que las covarianzas cruzadas $C_{12}(h) = C_{21}(h)$ del sistema de cokrigado pueden escribirse en función de la covarianza de los datos duros como:

$$C_{12}(h) = \frac{C_{12}(0)}{C_1(0)} C_1(h), \quad \forall h$$

o lo que es lo mismo:

$$\rho_{12}(h) = \rho_{12}(0) \rho_1(h), \quad \forall h$$

donde $\rho_1(h)$ es el correlograma de la variable principal, $\rho_{12}(h)$ es el correlograma cruzado entre las variables principal y secundaria; y $\rho_{12}(0)$ es el coeficiente de correlación entre éstas.

Es decir que, bajo un modelo de Markov, la covarianza cruzada $C_{12}(h)$ se calcula reescalando la covarianza de la variable principal. El factor de escala es el cociente entre $C_{12}(0)$ y $C_1(0)$. El *estimador por cokrigado colocalizado bajo un modelo de Markov* en su forma estandarizada de la forma siguiente:

$$\frac{z_1^*(u_0) - m_1}{\sigma_1} = \sum_{\alpha=1}^{n_1} \lambda_{\alpha}^{(1)} \frac{z_1(u_{\alpha}) - m_1}{\sigma_1} + \lambda^2 \frac{z_2(u_0) - m_2}{\sigma_2}$$

donde los pesos son la solución del sistema siguiente:

$$\sum_{\beta=1}^{n_1} \lambda_{\beta}^{(1)} \rho_1(u_{\alpha} - u_{\beta}) + \lambda^{(2)} \rho_{12}(0) \rho_1(u_{\alpha} - u_0) =$$

$$= \rho_1(u_{\alpha} - u_0), \quad \alpha = 1, \dots, n_1$$

$$\sum_{\beta=1}^{n_1} \lambda_{\beta}^{(1)} \rho_{12}(0) \rho_{(1)}(u_0 - u_{\beta}) + \lambda^{(2)} \rho_2(0) = \rho_{12}(0)$$

En este sistema, si $\rho_{12}(0)=0$, entonces $\lambda^{(2)}=0$ y el dato secundario es ignorado. Si $\rho_{12}(0)=1$, el sistema se transforma en uno de krigado simple con $(n+1)$ datos, cuyo resultado es $\lambda^{(2)}=1$ y $\lambda_{\beta}^{(1)}=0$ para todo β , es decir el estimador es igual al valor de la variable secundaria.

El algoritmo de cokrigado colocalizado markoviano requiere el mismo esfuerzo de implementación que el de krigado con una deriva externa, y tiene la importante ventaja de permitir que el dato secundario participe directamente en la estimación de la variable principal. Además, y a diferencia del krigado con una deriva externa, el modelo de cokrigado colocalizado se construye en base a un parámetro de calibración como es el coeficiente de correlación, lo que hace que la elección de tal modelo pueda verificarse a partir de los datos. Sin embargo, el cokrigado colocalizado requiere que la variable secundaria sea conocida en todos los puntos donde se quiera estimar la variable principal.

Los algoritmos hasta aquí presentados dan como resultado un valor estimado de la variable principal en cada localización considerada. Este valor estimado se construye en función de los valores que las variables principal y secundaria toman en el entorno del punto a estimar, de forma diferente según la técnica considerada. Los algoritmos que presentamos a continuación no producen una estimación del valor de la variable modelizada, sino de la probabilidad de que estos valores no superen un dado umbral. Como veremos, esto permite inferir en cada localización la distribución completa de los valores de la variable. Estas distribuciones, que también se construyen considerando la información local y vecina de ambas variables, constituyen un modelo de incertidumbre del atributo de interés, a partir del cual es posible elegir un

valor estimado para aquellos nodos donde la variable principal se desconoce. Estos métodos trabajan con variables indicadoras las cuales son introducidas en el siguiente apartado.

5.7 Krigado indicador

La variable indicadora $I(u; z_k)$ asociada a la variable $Z(u)$ para el umbral z_k en el punto de coordenadas u , se define de acuerdo a la siguiente función:

$$I(u; z_k) = \begin{cases} 1 & \text{si } Z(u) \leq z_k \\ 0 & \text{si } Z(u) > z_k \end{cases}$$

En el krigado indicador (ver Journel, 1983, 1986, 1989; Alabert, 1987; Isaaks y Srivastava, 1989; y Gómez-Hernández, 1991, 1993a), el rango de la variable en estudio se discretiza en $K+1$ clases por medio de K umbrales. Por ejemplo, consideremos una variable cuyo rango de variación es discretizado en tres clases por medio de dos umbrales con valores 25 y 50 respectivamente. Sea un dato cuyo valor es 43 localizado en un punto de coordenadas u . Las funciones indicadoras tomarán los siguientes valores: $i(u; 25)=0$ e $i(u; 50)=1$. Si para cada dato se realiza este mismo razonamiento, la transformación de la variable original de acuerdo al formalismo indicador, da como resultado un grupo de vectores, uno para cada localización u donde hay un dato disponible, integrado por K miembros $i(u; z_k)$, $k=1, \dots, K$.

Supongamos que realizamos un krigado ordinario de las variables indicadoras para cada umbral de manera que obtenemos el siguiente estimador en un punto de coordenadas u_0 :

$$i^*(u_0; z_k) = \sum_{\alpha=1}^n \lambda_{\alpha}(u_{\alpha}; z_k) \cdot i(u_{\alpha}; z_k)$$

Para obtener los coeficientes λ_{α} , que en el caso más general son función tanto de las coordenadas como del umbral, es necesario resolver el siguiente sistema de ecuaciones:

$$\begin{aligned} \sum_{\beta=1}^n \lambda_{\beta}(u_{\beta}; z_k) C_I(u_{\alpha} - u_{\beta}; z_k) + \mu(u_{\beta}; z_k) &= \\ = C_I(u_{\alpha} - u_0; z_k), & \quad \alpha = 1, \dots, n \end{aligned}$$

$$\sum_{\beta=1}^n \lambda_{\beta}(u_{\beta}; z_k) = 1$$

donde $C_i(h; z_k)$, con h igual a la separación entre puntos, es la función de covarianza de la variable indicadora, en el caso más general diferente para cada umbral considerado, y μ es el parámetro de Lagrange. El concepto de covarianza indicadora es idéntico al de covarianza clásica, sólo que calculada para las variables indicadoras.

El valor $i^*(u_0; z_k)$, generalmente comprendido entre 0 y 1, es como tal un mal estimador de la variable indicadora, puesto que ésta sólo puede tomar los valores 0 y 1. Pero cuando este valor se interpreta como un estimador del valor esperado de la variable indicadora $E\{i(u_0; z_k)\}$, resulta que $i^*(u_0; z_k)$ es el mejor estimador lineal de la probabilidad condicional acumulada de que la variable esté por debajo del umbral. Esto es consecuencia de la misma definición de variable indicadora cuyo valor esperado es igual a la probabilidad acumulada. De esta manera:

$$\begin{aligned} F[u_0; z_k | (n)]^* &= \text{Prob}[Z(u_0) \leq z_k | (n)]^* = \\ &= \sum_{\alpha=1}^n \lambda_{\alpha}(u_{\alpha}; z_k) i(u_{\alpha}; z_k) \end{aligned}$$

donde $F[\cdot]^*$ es la función de distribución de probabilidad acumulada condicionada a los n datos que intervienen en la estimación. El krigado de la variable indicadora para todos los umbrales z_k , $k=1, \dots, K$, en cada punto u , proporciona una estimación de la función de probabilidad acumulada de la variable estudiada. Estrictamente hablando proporciona valores discretos de la función de probabilidades acumuladas (de acuerdo al número de umbrales seleccionados) que interpolados y extrapolados adecuadamente, dan como resultado una función continua. De esta función de probabilidad acumulada se puede seleccionar el valor medio o el valor mediano como estimador de la variable $z^*(u_0)$.

El krigado indicador tiene una ventaja importante sobre las técnicas de krigado clásicas, en cuanto a que es capaz de incluir información imprecisa o blanda. El formalismo indicador permite codificar de un modo general y sencillo los datos que provienen de fuentes diferentes y con distintos grados de confianza. Para comprender este punto veamos el siguiente ejemplo.

Supongamos que el rango de variación de la variable original que se intenta estimar es $[0, 100]$, por ejemplo la porosidad. Se ha decidido discretizar este rango en $K=10$ clases, por ejemplo:

$$\begin{aligned} \{z_k, k=1, \dots, 9\} &= \\ &= \{5, 10, 20, 30, 40, 60, 80, 90, 95\} \end{aligned}$$

Se cuenta con tres datos que provienen de fuentes distintas. Una medida de porosidad obtenida en laboratorio sobre un testigo extraído de una perforación que se puede considerar exacta, es decir un *dato duro* (incertidumbre nula o despreciable), y que vale $z=25$. Un dato calibrado por regresión lineal a partir de información geoelectrónica que se sabe pertenece al intervalo $[15, 35]$. Este tipo de dato se denomina *dato de intervalo*. Por último, en un punto donde no tenemos medida registrada, adoptamos una *distribución previa* obtenida empíricamente por el juicio de un experto, y definida por los siguientes cuantiles: $\{P(Z \leq z_k), k=1, \dots, 9\} = \{0.0, 0.0, 0.3, 0.5, 0.7, 1.0, 1.0, 1.0, 1.0\}$. Estos cuantiles expresan la probabilidad de que el valor de la porosidad se encuentre por debajo de un cierto umbral, por ejemplo la probabilidad de que la porosidad este por debajo del 20% se estima en 0.3.

Aplicando la definición de función indicadora cada dato se transforma en un vector de variables indicadoras como sigue:

a) dato duro

$z = 25 \rightarrow \{0, 0, 0, 1, 1, 1, 1, 1, 1\}$; es decir que un dato duro está representado por un vector completo de indicadores duros. El vector está completo pues conocemos el valor exacto del dato, y los indicadores se dicen duros pues sólo toman los valores 0 o 1.

b) dato de intervalo

$z \in [15, 35] \rightarrow \{0, 0, ?, ?, 1, 1, 1, 1, 1\}$; es decir que un dato de intervalo está representado por un vector incompleto de indicadores duros. El vector es incompleto ya que los indicadores pueden ser calculados fuera de los límites del intervalo pero no dentro de éste.

c) función de probabilidades previa

$\{0.0, 0.0, 0.3, 0.5, 0.7, 1.0, 1.0, 1.0, 1.0\}$. Este tipo de datos no se ve afectado por la transformación indicadora ya que fue definido de acuerdo a la probabilidad de estar por encima (o por debajo) de cada umbral. El vec-

tor resultante está completo e integrado por indicadores "blandos" con valores entre 0 y 1.

De esta manera, toda la información disponible queda codificada en vectores de indicadores que son integrados por el algoritmo de krigado indicador para la estimación de la porosidad. Una vez que todos los datos han sido codificados de acuerdo al formalismo indicador, no hay diferencia entre información dura y blanda desde un punto de vista algorítmico. Para el sistema de krigado indicador toda la información está en forma de vectores de indicadores.

En un modelo como el presentado arriba no es posible distinguir entre el patrón de continuidad espacial de los datos duros y de los blandos, ya que el estimador construido es uno krigado. En algunas aplicaciones es importante que pueda distinguirse entre los patrones de continuidad de los indicadores correspondientes a datos duros y los de los correspondientes a datos blandos. Para el ejemplo que hemos sugerido en este informe, dada la escasez de los datos de pozos es poco realista inferir las variaciones de la porosidad en la región entre perforaciones a partir de esta única información. En la práctica es común contar con información secundaria tal como la derivada de una campaña de sísmica. Esta información puede ser incorporada con el objetivo de lograr una mejor estimación de la porosidad entre pozos, ya que es esperable una buena relación entre los patrones de continuidad de ambos parámetros. Para ello se plantea la necesidad de un algoritmo que, además de considerar los distintos tipos de datos, distinga entre sus patrones de continuidad. Una solución pasa por considerar un estimador cokrigado, tal como el siguiente:

$$i^*(u_0; z_k) = \sum_{\alpha=1}^{n_1} \lambda_{\alpha}^{(1)}(u_{\alpha}; z_k) i(u_{\alpha}; z_k) + \sum_{\alpha=1}^{n_2} \lambda_{\alpha}^{(2)}(u'_{\alpha}; z_k) i'(u'_{\alpha}; z_k)$$

En este estimador se distingue la influencia de los n_1 datos duros y los n_2 datos blandos codificados bajo el formalismo indicador.

La resolución del sistema de cokrigado en correspondencia con el estimador anterior, requiere el laborioso cálculo de, como mínimo, dos autocovarianzas y una covarianza cruzada para cada umbral si sólo una variable secundaria es considerada. Como dijimos al presentar el algoritmo de cokri-

gado clásico, la experiencia ha mostrado que ésta es la única razón por la que el cokrigado indicador no es muy usado en la práctica. Para salvar esta dificultad se presenta el siguiente algoritmo, el cual recurre a un modelo markoviano para aliviar el cálculo de las funciones de covarianza.

5.8 Cokrigado indicador bajo un modelo de Markov-Bayes

Como en la técnica de krigado de indicadores presentada en el apartado anterior, en el cokrigado indicador (ver Journel y Zhu, 1990; Zhu, 1991; Zhu y Journel, 1992), la información disponible, tanto dura como blanda, puede ser codificada en cada localización como distribuciones previas acumuladas de probabilidad del atributo modelizado. Estas distribuciones previas locales son luego mejoradas utilizando los datos duros y blandos cercanos, construyendo así distribuciones posteriores a partir de las cuales podemos seleccionar un valor como estimador de la variable estudiada. El algoritmo de Markov-Bayes es una técnica capaz de derivar dichas distribuciones posteriores, condicionadas tanto a la información dura como a la blanda.

Bajo un esquema bayesiano todos los datos, duros y blandos, pueden ser codificados como distribuciones locales previas. De forma general hemos dicho que esto puede escribirse como sigue:

$$y(u; z) = \text{Prob} \{Z(u) \leq | \text{información local}\}$$

donde $Z(u)$ es una función aleatoria representando la variable principal de interés; u representa las coordenadas espaciales; e $y(u; z)$ denota la distribución local previa; y es una probabilidad escalar entre 0 y 1 que especifica la información local disponible en u .

Un dato duro $z(u_1)$, caracterizado por un vector de indicadores duros completo, es tratado como una función de densidad acumulada con varianza nula, esto es:

$$i(u_1; z) = \begin{cases} 0 & \forall z < z(u_1) \\ 1 & \forall z \geq z(u_1) \end{cases}$$

Un dato de intervalo indicando que $z(u_2)$ pertenece al intervalo $[a_2, b_2]$ es codificado como un vector de indicadores duros pero incompleto, y es

considerado como una función de densidad acumulada incompleta, esto es:

$$y(u_2; z) = \begin{cases} 0 & \forall z < a_2 \\ \text{indefinida} & \forall z \in [a_2, b_2] \\ 1 & \forall z > b_2 \end{cases}$$

Donde hay un dato secundario $v(u_3)$ es necesario calibrar una función de distribución acumulada local que represente la distribución previa de la variable principal condicionada a la información secundaria local, esto es:

$$y(u_3; z) = \text{Prob} \{Z(u_3) \leq z \mid v(u_3)\} \in [0, 1]$$

Para realizar esta calibración se representa gráficamente un diagrama de dispersión de los datos de calibración tal como el que mostramos en la figura 2.

Se divide el rango de las variables en un número dado de clases, y se procede a calcular lo siguiente:

$$\text{proporción} \{Z \leq z_k \mid V \in v_l\}, \quad l = 1, \dots, 10$$

donde Z es la variable principal, V es la variable secundaria, z_k es el k -ésimo umbral de Z , y v_l el l -ésimo intervalo de V . Los cuantiles se calculan por simple conteo de la cantidad de medidas de Z por debajo del umbral considerado en cada uno de los intervalos de V . La tabla 1 muestra los resultados obtenidos para el diagrama de la figura 2.

Por último, en aquellos nodos donde no existe ningún tipo de información, se debe adoptar una función de distribución $F(z)$ denominada *global*, previa a cualquier estimación y común a todos los nodos, la cual puede estar relacionada con el rango de validez de la variable. Por ejemplo, una manera de definir de forma global la porosidad es diciendo que sus valores estarán comprendidos en el intervalo $(0, 100]$. Está claro que éste es un nivel mínimo de información puesto que no contamos con ningún tipo de dato.

En resumen, para cada localización u , las distribuciones previas locales $y(u; z)$ toman la forma de vectores de valores de probabilidad, con hasta K miembros $y(u; z_k)$, donde $k=1, \dots, K$, siendo K los valores de los umbrales elegidos para discretizar la variable principal. Es importante tener presente que estas distribuciones están condicionadas sólo a la información local.

Las distribuciones posteriores en cada localización son obtenidas considerando la información vecina, esto es, las distribuciones cercanas a cada localización. Esta mejora produce una distribución de probabilidades que en general será diferente a la previa y se escribe de forma general como sigue:

$$\text{Prob}\{Z(u) \leq z \mid \text{ toda la información disponible}\}$$

Las distribuciones previas locales $y(u; z)$ pueden ser interpretadas como realizaciones de una función aleatoria $Y(x; z)$. Luego el estimador cokrigado a ser utilizado para mejorar los valores de y a partir de sus valores vecinos se escribe:

$$\text{Prob}^* \{Z(u) \leq z \mid (n)\} = \lambda_0 F(z) + \sum_{\alpha=1}^n \lambda_{\alpha} i(u_{\alpha}; z) + \sum_{\beta=1}^{n_2+n_3} v_{\beta} y(u'_{\beta}; z)$$

donde $F(z)$ es la distribución previa global; $i(u_{\alpha}; z)$ son las distribuciones previas generadas por los n_1 datos duros en las localizaciones; eson las distribuciones previas generadas por los n_2 datos blandos en las localizaciones u'_{β} . Los pesos $\lambda_0, \lambda_{\alpha}$ y v_{β} surgen de resolver un sistema de ecuaciones lineales similar al presentado en el apartado dedicado al cokrigado tradicional.

Como en todo estimador cokrigado la regresión anterior requiere para cada umbral z la definición de tres modelos de covarianzas, si al menos una variable secundaria es incorporada, los correspondientes a las funciones $I(u; z)$ e $Y(u; z)$ y su covarianza cruzada. Una asunción de tipo markoviana, similar a la presentada al hablar de cokrigado colocalizado, reduce el cálculo de las covarianzas a la modelización de las correspondientes a los datos duros. La autocovarianza de la información blanda y la covarianza cruzada entre datos de distinto tipo se obtienen como una función de la covarianza de la información dura y unos pocos parámetros derivados del grupo de datos de calibración. Las expresiones resultantes son las siguientes:

$$C_N(h; z) = B(z) \cdot C_l(h; z), \quad \forall h$$

$$C_Y(h; z) = B^2(z) \cdot C_l(h; z), \quad \forall h > 0$$

$$C_Y(0; z) = V_c^2(z) + V_f^2(z)$$

Tabla 1
Funciones de distribución (acumuladas) de probabilidad de la información principal (Z) condicionada a la información secundaria local (V).

		Variable principal Z										
		umbral	0.31	6.09	16.62	46.22	134.5	240.8	350.3	549.5	1073.0	Z _{max}
Var. sec. V		30.16	0.70	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		102.74	0.30	0.80	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		171.19	0.00	0.30	0.65	0.95	0.95	1.00	1.00	1.00	1.00	1.00
		242.0	0.00	0.00	0.30	0.55	0.70	0.85	0.95	0.95	1.00	1.00
		326.38	0.00	0.00	0.10	0.30	0.65	0.80	0.80	0.90	0.95	1.00
		433.83	0.00	0.00	0.05	0.15	0.35	0.50	0.80	0.90	0.90	1.00
		530.52	0.00	0.00	0.00	0.05	0.25	0.55	0.80	0.85	0.90	1.00
		671.98	0.00	0.00	0.00	0.00	0.10	0.25	0.50	0.60	0.75	1.00
		783.19	0.00	0.00	0.00	0.00	0.00	0.05	0.15	0.50	0.80	1.00
		V _{max}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.30	0.70	1.00

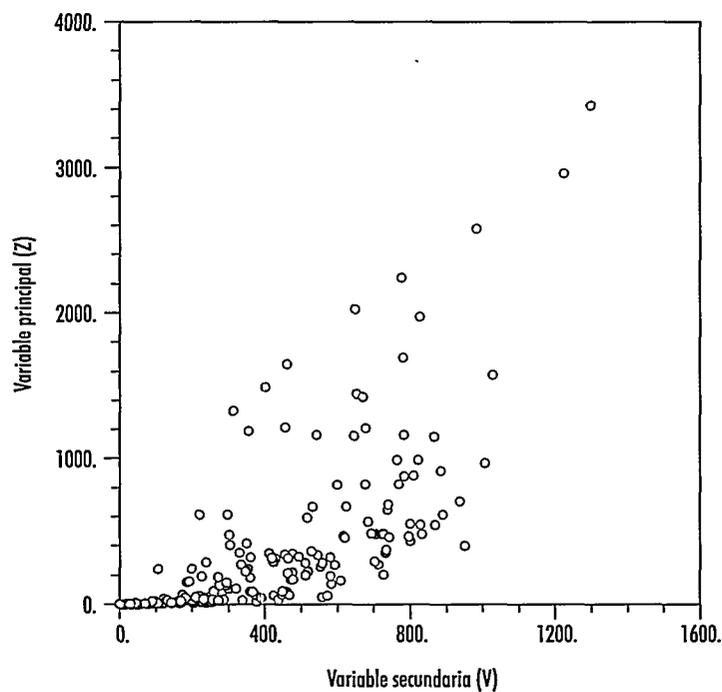


Figura 2. Diagrama de dispersión de 200 datos de calibración.

con:

$$C_I(h; z) = \text{Cov}\{I(u; z), I(u + h; z)\}$$

$$m^{(1)}(z) = E\{Y(u; z) \mid I(u; z) = 1\} \in [0, 1]$$

$$m^{(0)}(z) = E\{Y(u; z) \mid I(u; z) = 0\} \in [0, 1]$$

$$B(z) = m^{(1)}(z) - m^{(0)}(z) \in [-1, 1]$$

$$\sigma^{2(1)}(z) = \text{Var}\{Y(u; z) \mid I(u; z) = 1\}$$

$$\sigma^{2(0)}(z) = \text{Var}\{Y(u; z) \mid I(u; z) = 0\}$$

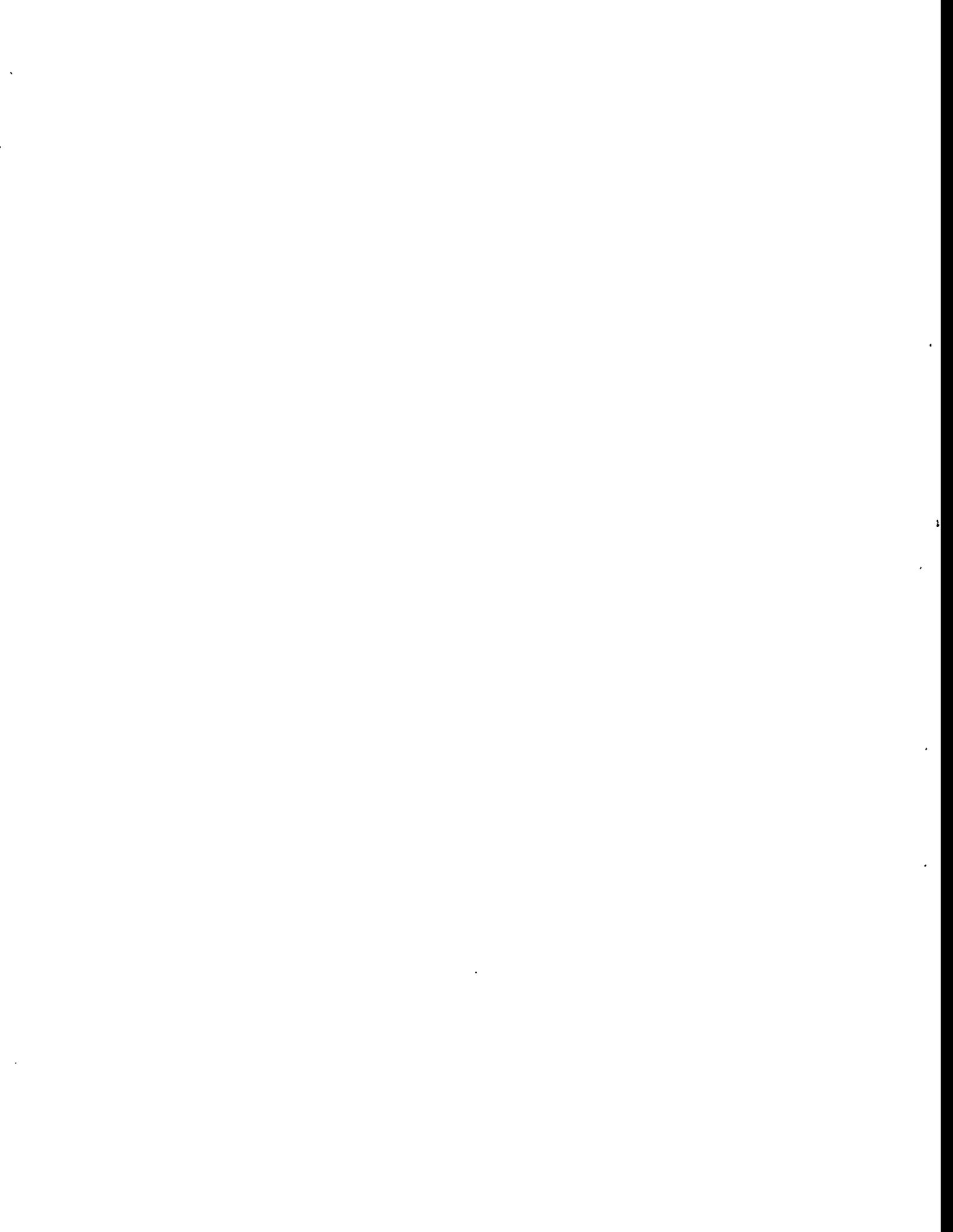
$$F(z) = \text{prob}\{Z(u) \leq z\} = E\{I(u; z)\}$$

$$V_c^2(z) = F(z)[1 - F(z)]B^2(z)$$

$$V_f^2(z) = F(z)\sigma^{2(1)}(z) + [1 - F(z)]\sigma^{2(0)}(z)$$

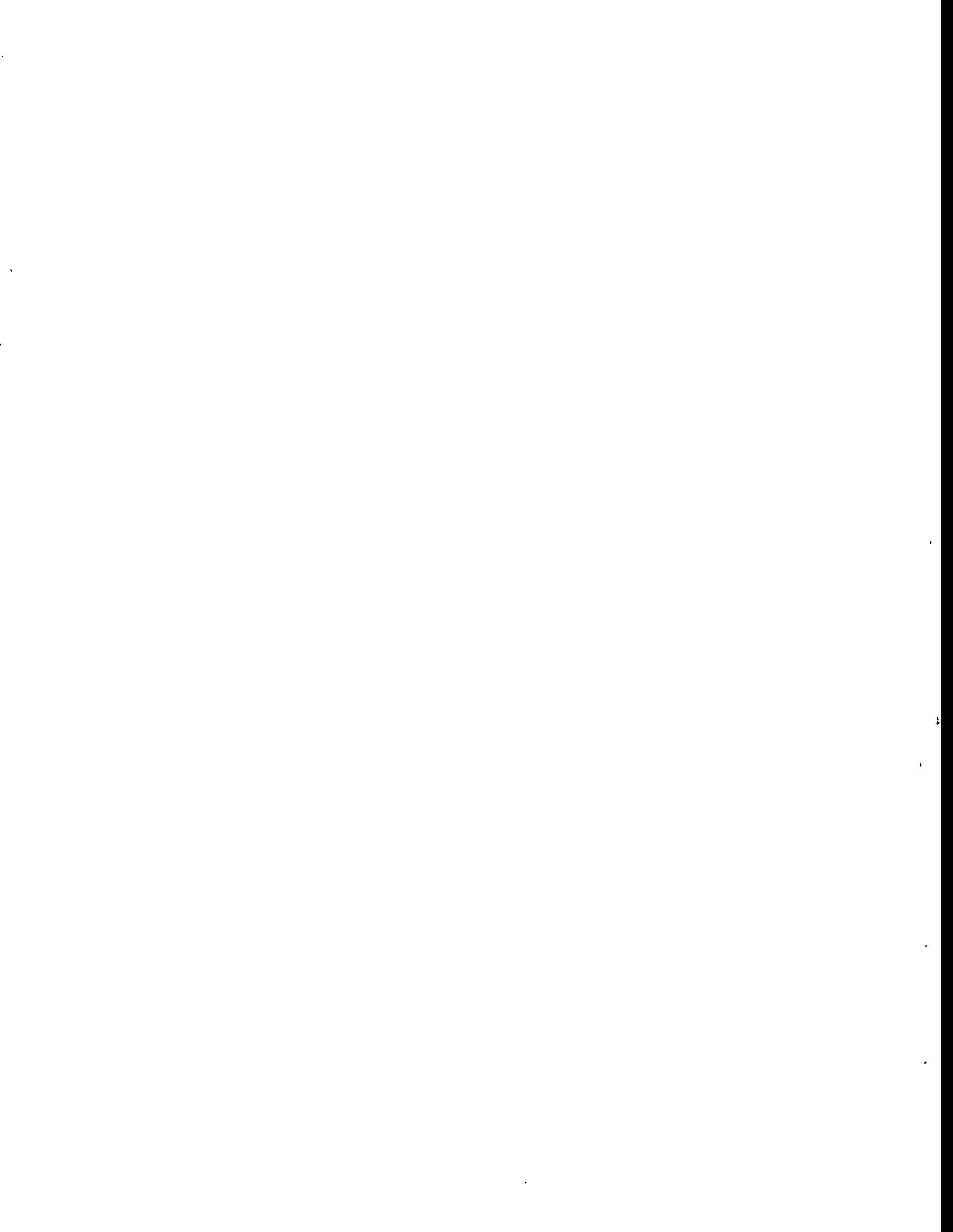
donde C_I , C_Y y C_{IY} son las autocovarianzas y covarianzas cruzadas respectivamente; h es el vector separación entre puntos considerados; z es un valor (umbral) de la variable principal no transformada Z ; y los restantes términos son parámetros obtenibles a partir de los datos de calibración.

El algoritmo de Markov-Bayes supone el mismo esfuerzo que el necesario para la técnica de krigado indicador, sin embargo tiene dos ventajas fundamentales sobre éste. Por un lado distingue las autocovarianzas de la información dura y blanda. Por el otro, a través del cokrigado de los indicadores, es posible mejorar las distribuciones previas locales calculadas a partir de la información blanda local, en función de la información vecina.



6

ALGORITMOS DE REPRESENTACION ESTOCASTICA



Los algoritmos de estimación dan como resultado una única representación del atributo de interés, esto es, la distribución espacial desconocida de un parámetro dado z es representada por un único grupo de valores estimados $\{z^*(u), u \in D\}$, siendo D el dominio del problema. Los métodos de interpolación anteriormente descritos, permiten obtener esos valores estimados en función de la información disponible acerca de la variable, pudiendo en algunos casos incorporar información secundaria. Bajo ciertas condiciones, estos métodos también proporcionan una medida de incertidumbre local. Sin embargo los campos estimados, dados los criterios con que son construidos, no reproducen la variabilidad real del fenómeno representado. Este efecto suavizante, propio de las técnicas de interpolación, puede resultar un problema en ciertas situaciones prácticas de la ingeniería. Para solucionar esta cuestión es necesario considerar técnicas que produzcan campos con un grado de variabilidad espacial de los parámetros de interés similar al observado. Estas técnicas se engloban bajo el concepto de *simulación estocástica*.

Se llama simulación estocástica al proceso de generar múltiples campos de la variable z modelizada como una función aleatoria $Z(u)$. Los campos simulados reproducen toda la información disponible y el patrón de continuidad espacial utilizada para la modelización de $Z(u)$.

6.1 Simulación secuencial

Uno de los algoritmos que permiten la generación de realizaciones múltiples es el de *simulación secuencial* (ver Gómez-Hernández y Cassiraga, 1994), cuyo esquema puede resumirse de la siguiente forma:

- 1) Se define un camino aleatorio visitando todos los nodos a ser simulados.
- 2) Para el nodo a simular determinar la función de distribución acumulada, condicionada a toda la información disponible, incluyendo los distintos tipos de datos y los valores previamente simulados en otros nodos. Conforme el proceso de la simulación se desarrolla, el número de valores presimulados a considerar en el cálculo de la función de distribución crece, con el consiguiente incremento en el tiempo de cálculo. Por esta razón, la función de distribución se condiciona sólo a los datos y va-

lores ya simulados más cercanos al nodo a simular.

- 3) Se selecciona aleatoriamente de acuerdo a la función de distribución obtenida en (2), un valor de la variable el cual se agrega como información condicionante para la simulación del próximo nodo.
- 4) Se retorna al paso (2) hasta que todos los nodos hayan sido simulados.
- 5) Este proceso se repite partiendo de otro camino aleatorio para generar otra realización.

La determinación de las funciones de distribución condicionadas puede hacerse paramétrica o no paramétricamente. En el primer caso, la función de distribución condicionada corresponde a un modelo conocido del que hay que determinar los parámetros, por ejemplo, en el caso de que la función de distribución se modelice como multigausiana, las funciones de distribución condicionales son gaussianas y su cálculo requiere el conocimiento de la media y la varianza. En el segundo caso, la función de distribución se determina a partir del cálculo de unos cuantos puntos de la misma. La simulación no paramétrica utiliza variables indicadoras, y se muestra como una técnica mucho más potente y flexible que la paramétrica.

6.1.1 Simulación gaussiana secuencial

Si adoptamos un modelo multigausiano para describir la dependencia espacial de las variables consideradas, se puede demostrar que las funciones de distribución condicionales a las que se hace referencia en el punto (2) del párrafo anterior son gaussianas, y por tanto es suficiente conocer la media y la varianza condicionales para su determinación.

De los métodos de simulación gaussiana secuencial describiremos los que pueden incorporar información secundaria, como son los basados en el cokrigado completo y en el cokrigado colocalizado con un modelo de Markov.

Puesto que los parámetros de interés no están en general normalmente distribuidos, la simulación gaussiana utiliza, en el caso de dos variables aleatorias, las variables transformadas siguientes:

$$Y_1(u) = \phi_1[Z_1(u)]$$

$$Y_2(u) = \phi_2[Z_2(u)]$$

donde Z_1 y Z_2 son las variables principal y secundaria respectivamente; u el vector de coordenadas espaciales; ϕ_1 y ϕ_2 representan las funciones que transforman Z_1 y Z_2 en las variables Y_1 e Y_2 con distribución normal de media nula y desviación típica unitaria. Estas funciones se determinan empíricamente a partir de los datos (Journel y Huijbregts, 1978, p. 478). Una vez finalizada la simulación se realiza la transformación inversa y se presentan los resultados en las variables originales.

6.1.1.1 Simulación gaussiana secuencial por cokrigado completo

El cokrigado completo de $Y_1(u)$ considerando tanto los datos primarios $y_1(u_\alpha)$, $\alpha=1, \dots, n_1$; como los secundarios $y_2(u'_\alpha)$, $\alpha=1, \dots, n_2$, proporciona la media y la varianza de la distribución condicional gaussiana de $Y_1(u)$. La media viene dada por el estimador cokrigado simple $y_1^*(u)$, y la varianza corresponde a la varianza del cokrigado simple $\sigma_{CKS}^2(u)$, esto es:

$$y_1^*(u) = \sum_{\alpha=1}^{n_1} \lambda_\alpha^{(1)} y_1(u_\alpha) + \sum_{\alpha=1}^{n_2} \lambda_\alpha^{(2)} y_2(u'_\alpha)$$

$$\sigma_{CKS}^2(u) = 1 - \sum_{\alpha=1}^{n_1} \lambda_\alpha^{(1)} S_1(u_\alpha - u) - \sum_{\alpha=1}^{n_2} \lambda_\alpha^{(2)} S_{12}(u'_\alpha - u)$$

donde $S_1(u_\alpha - u)$ es la covarianza de $Y_1(u)$; $S_{12}(u'_\alpha - u)$ es la covarianza cruzada entre $Y_1(u)$ y $Y_2(u)$. Los (n_1+n_2) coeficientes λ_α se obtienen resolviendo el sistema de cokrigado simple correspondiente (ver sección 5.4 de este informe). Un estimador como el anterior requiere modelizar las autocovarianzas y covarianzas cruzadas entre variables, y resolver un sistema de (n_1+n_2) ecuaciones para cada nodo simulado. La consideración del cokrigado colocalizado bajo un modelo markoviano alivia la carga de cálculo.

6.1.1.2 Simulación gaussiana secuencial por cokrigado colocalizado bajo un modelo markoviano

Esta metodología es idéntica a la anterior excepto en que sólo se considera el dato secundario colocalizado para realizar cada estimación, y en que una hipótesis de tipo markoviana reduce el cálculo de las covarianzas necesarias, tal como fue expli-

cado en el apartado 5.6 dedicado a cokrigado colocalizado bajo un modelo de Markov.

La media y la varianza de las funciones de distribución acumuladas condicionales gaussianas vienen dadas por las siguientes expresiones:

$$y_1^*(u) = \sum_{\alpha=1}^{n_1} \lambda_\alpha^{(1)} y_1(u_\alpha) + \lambda^{(2)} y_2(u')$$

$$\sigma_{CKS}^2(u) = 1 - \sum_{\alpha=1}^{n_1} \lambda_\alpha^{(1)} S_1(u_\alpha - u) - \lambda^{(2)} S_{12}(0)$$

Esta estimación requiere la resolución de un sistema de (n_1+1) ecuaciones, y la estimación de la función de covarianza de la variable principal. Las funciones de covarianza secundaria y cruzada se obtienen por simple reescalado de la correspondiente a la variable principal como se explica en el apartado 5.6. En Journel *et al.* (1992) la simulación gaussiana bajo un modelo de Markov fue aplicada al cartografiado de un domo de sal en el Golfo de México.

6.1.1.3 Simulación conjunta de varias variables

Una extensión de las metodologías previas es la cosimulación secuencial, en la cual se simulan varias variables conjuntamente. Estas variables están por lo general correlacionadas entre sí, por ejemplo: la permeabilidad horizontal, la permeabilidad vertical y la porosidad de un acuífero. La cosimulación gaussiana secuencial considerando un cokrigado completo está descrita por Gómez-Hernández y Journel (1993) y por Verly (1993). La técnica es idéntica a la desarrollada para una variable, sólo que en lugar de simular un valor en cada nodo, se simula un vector de q componentes, donde $q=1, \dots, Q$; es el número de variables principales. Cada nodo simulado se condiciona a los datos disponibles más cercanos de la variable a simular, de las demás variables principales, de las variables secundarias, y a los valores previamente simulados. Esto, más la inferencia de las autocovarianzas y covarianzas cruzadas entre todas las variables, supone un mayor costo computacional con respecto a la simulación secuencial de una sola variable.

Almeida (1993a y b) y Almeida y Journel (1994) han desarrollado el algoritmo de cosimulación secuencial con información secundaria colocalizada y bajo un modelo de Markov. Este modelo reduce el número de datos secundarios participantes en

cada estimación al dato colocalizado para cada variable secundaria incorporada, y para cada variable principal previamente simulada. Además una hipótesis markoviana reduce sensiblemente el número de covarianzas a estimar, ya que tanto las autocovarianzas como las covarianzas cruzadas para las variables secundarias pueden expresarse como una función de las covarianzas de las variables principales. El proceso de cosimulación requiere establecer una jerarquía de atributos, tanto principales como secundarias, la cual establece el orden en que las variables son simuladas. En Almeida *et al.* (1993) y en Frykman (1994), la cosimulación gaussiana bajo un modelo de Markov es aplicada a un caso real para cartografiar las propiedades petrofísicas de un yacimiento en el Mar del Norte (Dan Field, Danish North Sea).

6.1.2 Simulación indicadora secuencial

La simulación indicadora secuencial no adopta ningún modelo estadístico a priori al contrario que la simulación gaussiana secuencial que adoptaba un modelo multigaussiano. La secuencia de la simulación comienza codificando toda la información disponible, incluyendo los datos blandos, de acuerdo a la función indicadora. Como se indicó en 5.7, el krigado indicador en sus distintas variantes proporciona un estimador no paramétrico de la función de distribución condicional acumulada que podemos utilizar directamente en el algoritmo de la simulación secuencial.

Tanto el krigado indicador como el cokrigado bajo un formalismo de Markov-Bayes, pueden utilizarse en la simulación secuencial. La diferencia entre un método y otro está en cómo es considerada la información blanda por el sistema de krigado/cokrigado. En el primer caso, y como fue dicho al hablar de krigado indicador, no hay posibilidad de distinguir los patrones de continuidad espacial de la información blanda de los de la información dura. En el caso del mejorado por Markov-Bayes los datos son incorporados a través de un estimador cokrigado, lo que además de permitir distinguir entre una variable y la otra, permite el mejorado de las funciones estimadas a priori, a partir de la información blanda, condicionándolas a los datos vecinos.

Chu *et al.* (1991, 1992) realizaron la simulación de la porosidad (a partir de datos de un reservorio de petróleo localizado en West Texas) aplicando tanto simulación gaussiana como simulación de indicadores bajo el algoritmo de Markov-Bayes.

La simulación gaussiana no puede incorporar información secundaria por lo que la simulación se realizó considerando sólo los datos duros, mientras que la simulación de indicadores incorporó la densa información secundaria disponible. Los campos generados por simulación de indicadores muestran una mayor resolución en cuanto a que son más estructurados y más consistentes con la información secundaria considerada, sobre todo se aprecia una mejor reproducción de los valores altos. Araktingi *et al.* (1992) cartografiaron un horizonte estratigráfico comparando los resultados entre los algoritmos de krigado por deriva externa y simulación secuencial indicadora bajo Markov-Bayes. Langlais y Doyle (1993) adaptaron el algoritmo de simulación indicadora secuencial para incorporar derivas.

6.2 Simulación por recocido simulado (annealing)

Los métodos indicadores permiten la incorporación directa de información proveniente de diferentes fuentes y bajo distintos formatos, siendo una de las técnicas más flexibles para la modelación estocástica en las ciencias del terreno. Sin embargo, los algoritmos indicadores son computacionalmente intensivos, requieren la inferencia de muchas funciones de covarianza, y no pueden considerar propiedades que no puedan caracterizarse mediante una función de covarianza.

La simulación por recocido simulado (ver Deutsch, 1992; Deutsch y Journel, 1991 y 1992b; Deutsch y Cockerman, 1994) es una herramienta capaz de complementar las técnicas más tradicionales de simulación estocástica. Su mayor ventaja es su capacidad de imponer ciertos patrones de variabilidad espacial que son difíciles de caracterizar a través de funciones de covarianza, a la vez que su facilidad, al menos teórica, para incorporar información proveniente de diversas fuentes. Su flexibilidad es posible gracias a un algoritmo simple pero computacionalmente poco eficiente por lo que el uso del recocido simulado no se recomienda para aquellos casos en que existan herramientas alternativas capaces de alcanzar el mismo objetivo.

El recocido simulado se basa en un algoritmo inicialmente desarrollado para la solución de problemas de optimización, en los cuales se requería la ordenación óptima de un sistema con un gran

número de componentes. En el contexto de la modelización estocástica los componentes podrían ser los valores del parámetro en estudio discretizados sobre una malla. La optimalidad de la ordenación se puede medir por cómo el orden de los bloques reproduce el patrón de correlación espacial (covarianzas) que se desea imponer. Puesto que existen muchas ordenaciones capaces de reproducir dicho patrón, encontrar un óptimo equivale a generar una realización del atributo modelizado. Así, por un lado el problema de representación estocástica es transformado en uno de optimización, y por el otro el problema de optimización es resuelto por recocido simulado, técnica que pertenece a las llamadas de relajación estocástica. La característica principal de los métodos de relajación estocástica es perturbar iterativamente la realización "candidata" y luego aceptar o no la perturbación en función de alguna regla de decisión. Esta regla está en relación con cuán cercanas las propiedades estadísticas de la representación perturbada están de las deseadas.

Para ilustrar la aplicación de esta metodología, se considera una variable categórica $Z(u)$ que puede tomar k valores, con $k=1, \dots, K$. El problema es generar una distribución espacial de $Z(u)$ en los N nodos de una malla, es decir simular los valores $z(u_i)$, $i=1, \dots, N$, preservando una cierta correlación espacial, que, por ejemplo, podemos suponer que viene dada por el histograma bivariado siguiente:

$$f(h; k_1, k_2) = \text{Prob}\{Z(u) \in \text{a la categoría } k_1, \\ Z(u+h) \in \text{a la categoría } k_2\}, k_1, k_2 = 1, \dots, K$$

Este histograma es conocido para los vectores de separación $h=h_1, \dots, h_L$. Una posible función objetivo puede ser:

$$O = \sum_{l=1}^L \sum_{k_1}^K \sum_{k_2}^K [f(h_l; k_1, k_2)_{\text{referencia}} - f(h_l; k_1, k_2)_{\text{realización}}]^2$$

El algoritmo general de recocido simulado puede esquematizarse a través de los siguientes pasos:

- 1) Establecer las componentes de referencia de la función objetivo, es decir, cuáles son los valores del histograma bivariado que queremos reproducir:

$$f(h_l; k_1, k_2)_{\text{referencia}, u}, l = 1, \dots, L, u; k_1 = 1, \dots, K, u; \\ k_2 = 1, \dots, K, u$$

- 2) Generar una realización inicial $z(u_i)$, $i=1, \dots, N$ que al menos reproduzca las frecuencias relativas de cada una de las categorías a las que puede pertenecer $Z(u)$. Por ejemplo generando los valores aleatoriamente a partir de la distribución univariada.
- 3) Calcular las componentes de la realización en la función objetivo, es decir:

$$f(h_l; k_1, k_2)_{\text{referencia}}, l = 1, \dots, L; k_1 = 1, \dots, K; k_2 = 1, \dots, K$$

- 4) Evaluar la función objetivo O .
- 5) Seleccionar dos nodos al azar e intercambiar sus valores. Este intercambio resulta en una nueva realización $z'(u_i)$, $i=1, \dots, N$.
- 6) Actualizar todas las componentes de la función objetivo para la nueva realización y reevaluarla, lo que da como resultado O' .
- 7) El intercambio (y por tanto la nueva realización) será aceptada si la función objetivo decrece, es decir, si $O' < O$ ó, con una cierta probabilidad, si la función objetivo aumenta. La segunda posibilidad de aceptación hace que una fracción de los intercambios que no ayudan a reducir la función objetivo se acepten evitando así que el proceso de optimización se quede estancado en un mínimo local. Esta probabilidad va decreciendo a medida que la simulación progresa de manera que cada vez es menor la fracción de tales intercambios que se aceptan. Es precisamente esta posibilidad de aceptar intercambios que no ayudan a reducir la función objetivo lo que le da el nombre de recocido simulado.
- 8) Si la función objetivo alcanza un valor cercano a cero, la realización se considera definitiva. De lo contrario se vuelve al paso (5) y se realiza un nuevo intercambio.

Este algoritmo puede sufrir muchas variantes según se consideren distintas funciones objetivos, diferentes modos de crear la representación inicial, diferentes mecanismos de perturbación, y distintas reglas de aceptación.

La técnica del recocido simulado anteriormente descrita puede adaptarse para considerar información blanda o secundaria. Para ello es necesario modificar la función objetivo de manera que pueda considerar esa información en el proceso de optimización. Por ejemplo, si se cuenta con un grupo de datos de calibración, es posible obtener

información acerca de la correlación entre las variable principal y secundaria, a partir del correspondiente diagrama de dispersión. Si sumamos esta correlación por el coeficiente de correlación lineal $\rho_{\text{calibración}}$ y queremos que la realización presente dicho coeficiente de correlación, no tenemos más que añadir un término a la función objetivo como el siguiente:

$$O_c = [\rho_{\text{calibración}} - \rho_{\text{realización}}]^2$$

De forma similar puede considerarse otro tipo de información mientras pueda expresarse su relación con la variable principal de una manera capaz de ser integrada en la función objetivo. Sin embargo, el diseño de una función objetivo compleja puede hacer impracticable el método desde el punto de vista computacional. En Deutsch y Journal (1992b) se presentó el esquema de una aproximación por recocido simulado para incluir permeabilidades efectivas derivadas de ensayos de bombeo en la simulación de campos de permeabilidad a escala local.

La mayor desventaja de esta técnica de simulación es que al tratarse de un problema de optimización, la convergencia hacia el óptimo puede ser lenta, reduciendo su eficiencia computacional. Existe la posibilidad de implementar métodos que aceleran esta convergencia, pero que pueden sesgar los resultados. Sin embargo, si partimos de una buena condición inicial, este sesgo puede ser mínimo.

6.3 Simulación por campo de probabilidades

Froidevaux (1992) y Srivastava (1992 y 1994) han presentado una técnica de simulación condicional capaz de incorporar diferentes tipos de información, conocida como *simulación por campo de probabilidades*. Dos son las características distintivas de este algoritmo. La primera es la disociación entre las tareas de estimar la función de distribución acumulada condicionada en cada localización y la generación de las representaciones equiprobables del atributo. (Recordemos que en la técnica de simulación secuencial las funciones de distribución se van calculando secuencialmente en cada punto a simular). La segunda es que la integración de distintos tipos de información se hace de un modo muy sencillo.

Consideremos un atributo Z a simular sobre los nodos de una malla, condicionado a una serie de medidas. Previo a la simulación podemos determinar en cada nodo la función de distribución condicional en base a un cierto modelo de función aleatoria por alguno de los métodos descritos en secciones anteriores. En concreto, en el nodo n conoceremos:

$$F(u; z | (n)) = \text{Prob}\{Z(u) \leq z | (n)\}$$

donde (n) representa la información condicionante local.

Ahora consideremos una realización del atributo $\{z_s(u_i), i=1, \dots, N\}$. Puesto que las funciones de distribución local son biunívocas hay una correspondencia inequívoca entre los valores de la realización y las correspondientes probabilidades asociadas a través de dichas distribuciones locales:

$$\{z_s(u_i), i=1, \dots, N\} \Leftrightarrow \{p_s(u_i), i=1, \dots, N\}$$

con $p_s(u_i) = F(u; z_s(u_i) | (n))$, donde las probabilidades locales $p_s(u)$ pueden ser interpretadas como los resultados de una función aleatoria $P(u)$. Esto plantea la posibilidad de generar una realización de las probabilidades a partir de las cuales, y a través de las funciones de distribución locales, obtener los valores del atributo.

Las principales ventajas de este procedimiento serían el que las realizaciones de los campos de probabilidad no tienen que ser condicionales y el que la función $F(u; z | (n))$ puede determinarse con los métodos de estimación más elaborados de que dispongamos.

La generación de los campos de probabilidad se realiza bajo las siguientes suposiciones:

- Dado que $P(u)$ representa un campo de probabilidades, este sigue una distribución uniforme; y
- Hay unas "características de permanencia" entre la transformada uniforme $U(Z(u))$ y $P(u)$, que nos permite asimilar la covarianza del campo de probabilidades al de la transformada uniforme de los datos disponibles, es decir:

$$C_p(h) \approx C_u(h)$$

con:

$$U(u) = \text{Prob}\{Z < z(u)\}$$

Este punto es el más intuitivo del algoritmo y no existe una justificación teórica del mismo. En Froidevaux (1992) estas hipótesis son contrastadas experimentalmente.

La implementación práctica del algoritmo de simulación de campo de probabilidades puede esquematizarse como sigue:

- 1) Definimos un mallado que cubra el área de interés.
- 2) Obtenemos para cada nodo de la malla la función de distribución acumulada condicionada local $F(u,z)$ del atributo a ser simulado.
- 3) Calculamos y modelizamos la función de covarianza de la transformada uniforme de $Z(u)$ y asumimos que:

$$C_p(h) \approx C_u(h)$$

- 4) Generamos sobre la malla una simulación no condicionada de $P(u)$ con distribución uniforme y covarianza $C_p(h)$.
- 5) Para cada nodo obtenemos el valor $z_s(u)$ a partir de la función de distribución local $F(u,z)$ utilizando el valor de probabilidad $p_s(u)$ correspondiente, es decir:

$$z_s(u) = F^{-1}(u, p_s(u))$$

- 6) Repetimos (4) y (5) para obtener otra realización equiprobable del atributo.

Bourgault *et al.* (1995) y Bourgault y Journel (1995) han extendido la técnica de simulación de campos de probabilidades para el caso de estimación de un campo de la variable principal en presencia de información secundaria densamente muestreada. Estos autores realizaron el cartografiado de la conductividad eléctrica del suelo incorporando medidas de respuesta electromagnética del mismo.

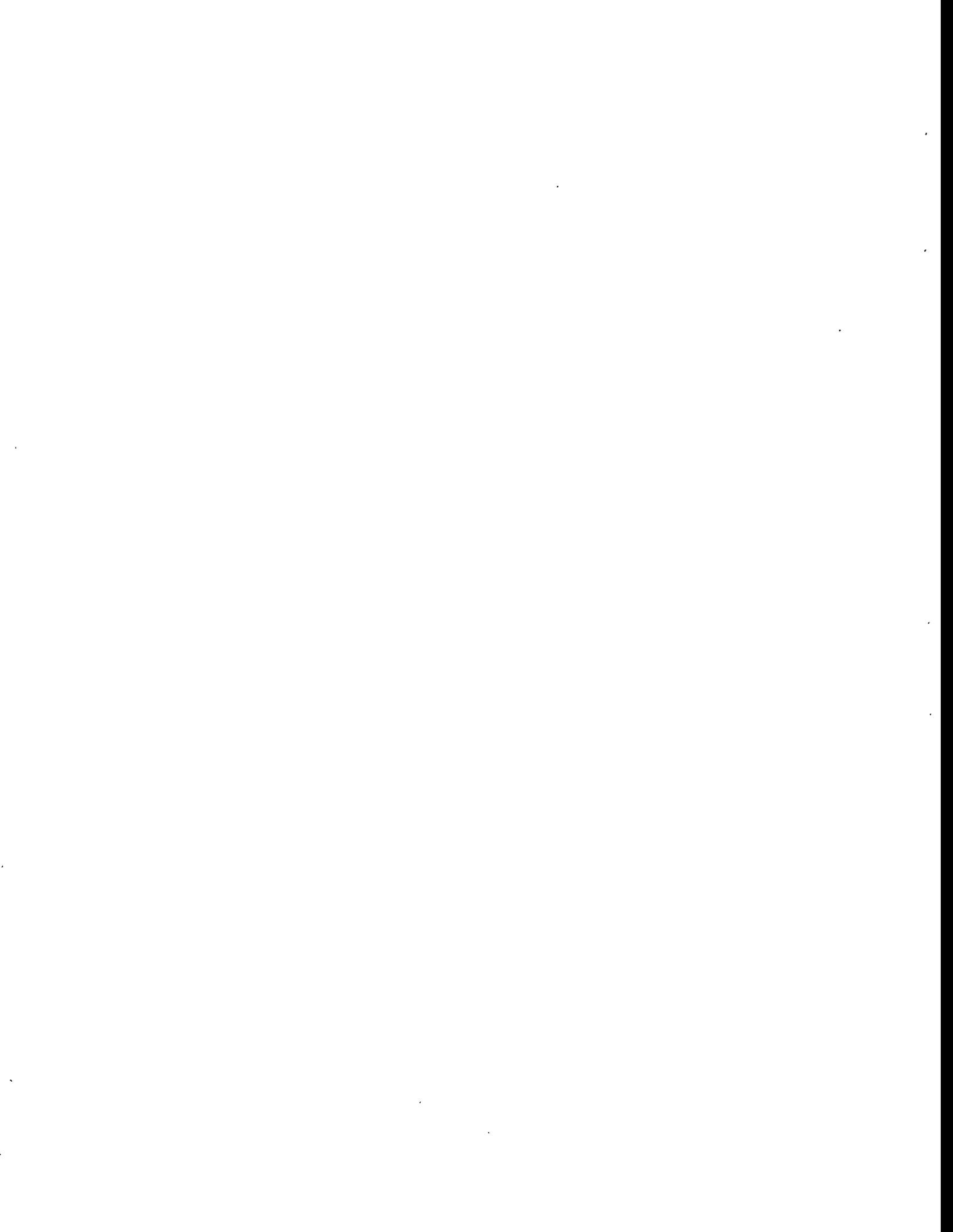
6.4 Algoritmos booleanos

Los algoritmos booleanos constituyen otra categoría dentro de los algoritmos de simulación estocástica. El objetivo de estos algoritmos es rellenar el espacio a simular con objetos geométricos en los que los parámetros que definen la geometría de estos objetos siguen una cierta ley de distribución. Un ejemplo clásico de modelo booleano es el de redes de fracturas para la modelización del flujo en macizos graníticos. Los objetos son fracturas planas elípticas y los parámetros geométricos que les definen son el centro, la orientación del semieje mayor y las longitudes de los semiejes mayor y menor.

Los modelos booleanos aunque atractivos por su fundamento geométrico son complejos de implementación (es difícil inferir las distribuciones de probabilidad de los parámetros que definen la geometría de los objetos) y muy difíciles de condicionar a la información existente.

7

EJEMPLO DE APLICACION



Para mostrar el resultado de las técnicas descritas en este informe presentamos un ejemplo sintético desarrollado sobre un acuífero rectangular de 2,0 por 1,2 kilómetros discretizado en 24000 celdas. Contamos con una base de datos de génesis natural obtenida a partir de un modelo de elevación digital de una zona de Nevada (Estados Unidos) denominada Lago Walker (Gómez-Hernández, 1992). Para obtener estos datos no se ha empleado ningún algoritmo de generación que pueda introducir algún artificio en su estructura de correlación, por lo que se puede afirmar que estos datos tienen un patrón de continuidad espacial susceptible de ser encontrado en la naturaleza.

La base de datos Walker cuenta con dos variables, U y V , exhaustivamente conocidas que constituyen los campos de referencia a partir de los cuales hemos seleccionado la información para desarrollar este ejercicio. La variable U , variable principal, representa permeabilidades (figura 3) y la variable V , variable secundaria (figura 4), un atributo geofísico correlacionado con aquélla, como por ejemplo la atenuación o la velocidad sísmica.

Como se aprecia en la figura 3, el campo de permeabilidades tiene una zona de valores altos rodeada por dos zonas de valores bajos. Se ha utilizado la misma escala de grises en ambas figuras para mostrar que, a pesar de que el patrón de continuidad espacial de ambos campos es similar, la variable secundaria tiene menos variabilidad que la variable principal, siendo el coeficiente de correlación lineal entre éstas igual a 0.6.

Del campo de la figura 3 se seleccionan 20 datos regularmente espaciados según el patrón de la figura 5, los cuales constituyen los datos duros. Estos datos corresponderían a medidas directas en perforaciones realizadas in situ. Por otro lado, consideramos que el campo de la figura 4 es el resultado de una campaña de prospección sísmica 3-D, lo que nos proporciona información exhaustiva acerca de la variable secundaria.

Para la estimación de los parámetros que cuantifican la correlación espacial de las variables en juego y su interrelación disponemos de un conjunto de datos de calibración, integrado por 200 parejas que fueron seleccionadas aleatoriamente de los campos exhaustivos de las figuras 3 y 4, y cuyas localizaciones pueden verse en la figura 6.

Este conjunto de datos no interviene en las estimaciones que de la variable principal se realizan más adelante, sólo es utilizado para obtener los parámetros de correlación necesarios por cada algoritmo. Algunos de estos parámetros se calibran

a partir de un diagrama de dispersión como el que se muestra en la figura 2, y que justamente corresponde a estos 200 datos de calibración.

En la figura 7 se presentan los variogramas obtenidos a partir de los datos de calibración para la variable principal y para la secundaria y el variograma cruzado entre éstas.

El variograma es una función vectorial que mide la distancia estructural entre dos puntos en función de la distancia euclídea. Es una función creciente que se estabiliza para un valor denominado meseta a una distancia llamada alcance. El alcance es la distancia a la cual la correlación de la variable desaparece, y en general, varía con la dirección. La pendiente del variograma es más tendida en la dirección de máxima continuidad indicando que la correlación se pierde más lentamente, y lo es menos en la dirección de mínima continuidad donde la correlación desaparece más rápidamente.

En nuestro ejercicio las direcciones de máxima y mínima continuidad son N300 y N60E respectivamente. En la figura 7 se muestran con línea de puntos el variograma experimental y en línea llena los modelos ajustados.

7.1 Resultados de los algoritmos de estimación

Estamos interesados en obtener un campo estimado de la variable principal U considerando toda la información disponible, esto es, los 20 datos duros y los 24000 datos blandos.

La solución más sencilla es despreciar los datos blandos e interpolar los 20 datos de la variable principal. En la figura 8 se muestra el resultado de realizar el krigeado simple de los 20 datos duros. El mapa resultante presenta una resolución pobre en comparación con el campo de referencia de la variable principal (figura 3). La interpolación realizada no considera la información secundaria, y sólo se ha realizado con el fin de comparar los resultados obtenidos cuando dicha información es incorporada en las estimaciones. La valoración de los campos estimados se hace comparándolos con el campo de referencia de la figura 3 y con el campo que resulta de interpolar la información principal sin incorporar la secundaria (figura 8).

7.1.1 Análisis de los resultados

Regresión lineal

A partir de los 20 datos de la variable principal y los correspondientes datos secundarios en las mismas localizaciones, se calibra una recta de regresión. Las estimaciones de la variable principal son obtenidas transformando el campo secundario de la figura 4 a través de dicha recta de regresión. Este método es simple y no es capaz de considerar la dependencia espacial entre las variables. Su bondad depende del grado de correlación lineal entre los datos. El campo resultante de la aplicación de este método se presenta en la figura 9. Como puede apreciarse corresponde a un escalado de la figura 4, y dado que disponemos de información secundaria densamente muestreada y que los patrones de continuidad espacial de las variables en juego son similares, la calidad del campo resultante es buena.

Krigeado con una deriva externa

Este método es aplicable cuando de alguna manera podemos afirmar que la variable secundaria define la tendencia de la variable principal. En nuestro caso el campo exhaustivo de las medidas de atenuación o velocidades sísmicas puede ser un buen informador de la tendencia de la porosidad (y de la transmisividad) del acuífero. El campo resultante se muestra en la figura 10. Como para el caso de regresión lineal este método requiere conocer la variable secundaria de forma exhaustiva, por lo que la calidad del campo interpolado se ve afectada si la información blanda es escasa. Además este método no explota la correlación espacial entre las variables consideradas.

Cokrigeado colocalizado

Este método incorpora la información secundaria como covariada, es decir que explota la información sobre continuidad espacial de la variable principal y la de la correlación lineal puntual entre la variable primaria y la secundaria. Al igual que los métodos anteriores requiere del conocimiento de la variable secundaria de forma exhaustiva. El campo resultante se muestra en la figura 11. Aunque, por lo menos a tenor de este ejercicio, no se puede afirmar que el resultado sea mejor que el de los métodos anteriores, la gran ventaja de este

algoritmo radica en su menor costo computacional, sobre todo si se desea incorporar más de una variable secundaria.

Cokrigeado completo

Toda la información es tratada como covariada, es decir que se utiliza la auto-correlación espacial de las variable principal y secundaria y la correlación cruzada entre las mismas. El campo resultante se muestra en la figura 12. De los métodos cuyos resultados hasta aquí hemos presentado, el cokrigeado completo es el que mayor información sobre la correlación entre las variables incorpora. En contrapartida requiere un mayor esfuerzo de modelización de variogramas y el tiempo de cómputo puede ser muy elevado si se quiere incorporar más de una variable secundaria. Es interesante comparar las imágenes 11 y 12 y apreciar la poca degradación que implica considerar sólo el dato secundario colocalizado.

Krigeado indicador

Como fue explicado en el punto 5.7, el krigeado indicador no produce una estimación del valor que la variable en juego puede tener, sino que estima valores de la probabilidad de que dicha variable se encuentre por debajo de determinado valor. Estos valores son interpolados y extrapolados obteniendo una función de distribución de probabilidades a partir de la cual se elige generalmente el valor medio o el mediano como estimador de la variable que interesa. En el ejemplo que aquí presentamos se ha seleccionado el valor mediano, siendo el campo resultante el que se muestra en la figura 13. El aspecto de este campo es el peor de los hasta aquí presentados ya que cada dato blando ha sido codificado como un dato de intervalo a partir del diagrama de dispersión de los datos de calibración siguiendo la metodología que comentamos a continuación. Se divide el rango de variación de la variable secundaria. Para nuestro ejercicio hemos considerado los deciles. Dentro de cada intervalo así definido para la variable secundaria se procede a obtener el valor mínimo y máximo que toma la variable principal. Estos valores definen el intervalo dentro del cual para un dado valor de la variable secundaria está el valor de la variable principal. Esto es, sabemos que si la variable secundaria toma un determinado valor, el valor de la variable principal estará entre un máximo y un mínimo pero sin saber exactamente cuál es ese valor.

Evidentemente cuanto menos dispersa sea la nube que forman los datos de calibración más estrechos serán los intervalos calibrados, lo que restringirá más la variabilidad de la variable principal y mejorará la calidad de la información incorporada. Una vez realizada esta calibración todos los datos secundarios son transformados de acuerdo a ella e incorporados en el proceso de krigado indistintamente con los datos duros. Esto supone una gran incertidumbre puesto que no conocemos cuanto vale la variable dentro de los límites de cada intervalo. El krigado indicador se comporta mejor que los métodos anteriores cuando la información secundaria no está densamente muestreada.

Cokrigado indicador

Al igual que el krigado indicador este método permite construir las distribuciones locales de frecuencia para la variable modelizada a partir de las cuales es posible seleccionar un valor de ésta. En este ejercicio hemos seleccionado el valor mediano. El campo resultante puede verse en la figura 14. Puede observarse que el resultado es muy bueno y apreciar el impacto que tiene el construir las distribuciones locales considerando los indicadores de las variables principal y secundaria por separado comparando esta figura con la anterior.

Campo de probabilidades

En la figura 15 puede verse el campo interpolado aplicando la extensión del algoritmo de simulación por campo de probabilidades a la que hicimos referencia en el punto 6.3. Si se dispone de la información secundaria de forma exhaustiva y de las funciones locales de distribución de probabilidades, posiblemente construidas a partir de alguno de los métodos presentados, el algoritmo es el más eficiente computacionalmente.

7.2 Resultados de los algoritmos de simulación

Para evaluar el comportamiento de los algoritmos de simulación hemos realizado un ejercicio en el que modelizamos el flujo de agua subterránea y el transporte de masa para el acuífero descrito en el punto 7.1. Los algoritmos evaluados fueron los siguientes:

- simulación gaussiana secuencial,
- simulación gaussiana secuencial por cokrigado colocalizado bajo un modelo markoviano,
- simulación indicadora secuencial,
- simulación secuencial por cokrigado indicador con un modelo de Markov-Bayes,
- simulación por campos de probabilidades, y
- simulación por recocido simulado.

Utilizaremos una aproximación de Monte Carlo, similar a la descrita por Gómez-Hernández y Wen (1994), para cuantificar la incertidumbre en la estimación del tiempo y de la posición de llegada de un contaminante que viaja a través del acuífero.

Primero, el problema de flujo y de transporte es resuelto para el campo de referencia (figura 16, campo superior). Por razones de eficiencia computacional el campo original de 200 por 120 celdas es escalado utilizando bloques de 2 x 2 celdas y asignando a éstos la media geométrica de los bloques individuales. En la figura 16 (campo inferior) se muestra el campo de referencia escalado. Las condiciones de borde que hemos considerado son: altura prescrita constante en los bordes oeste y este iguales a 10 y 0 metros respectivamente, y flujo nulo a través de los bordes norte y sur. El mapa de alturas piezométricas resultante se muestra en la figura 17. Puede observarse que a causa de las condiciones de borde el flujo medio se desarrolla en sentido oeste-este.

Para el problema de transporte hemos considerado el caso de un hipotético repositorio de residuos nucleares localizado en el borde oeste del acuífero. Un escape de 2000 partículas uniformemente repartidas en un segmento de 0,2 kilómetros de longitud, cuyo centro se encuentra en la coordenada vertical $y=0,6$ km, es modelizado considerando sólo el transporte advectivo. En las figuras 16 y 17 se muestran las trayectorias seguidas por 20 de esas partículas representadas sobre el campo de transmisividades escaladas y sobre el de alturas piezométricas, respectivamente. En la figura 18 hemos representado gráficamente la curva de llegada del contaminante mientras que en la 19 y 20 los histogramas de los tiempos y de las posiciones de llegada al borde este para 2000 partículas respectivamente. Las mismas figuras fueron obtenidas a partir de los resultados obtenidos por cada uno de los métodos utilizados con el fin de ser comparados con las correspondientes al campo de referencia.

Para el escenario descrito un análisis de la incertidumbre en la predicción de los tiempos y las posiciones de llegada es realizado utilizando los datos descritos en 7.1. La metodología seguida, la misma en cada uno de los métodos, es la siguiente:

- a) Generamos 100 simulaciones del campo de transmisividad condicionadas tanto a la información dura como a la blanda (excepto para el algoritmo de simulación gaussiana secuencial que no permite la incorporación de información blanda), y con los patrones de variabilidad espacial modelizados a partir de los datos de calibración.
- b) Cada uno de estos campos es escalado según bloques de 2 x 2 a los cuales asignamos la media geométrica de los bloques individuales.
- c) La ecuación del flujo subterráneo es resuelta para cada uno de los campos obtenidos en el paso anterior.
- d) Las trayectorias de las partículas son evaluadas en cada uno de los campos simulados utilizando el programa descrito en Wen y Kung (1993) considerando sólo el movimiento advectivo del contaminante.
- e) Los resultados son analizados estadísticamente. Este análisis consiste en obtener para el conjunto de 100 realizaciones de cada método, la curva media de llegada, las curvas envolventes correspondientes al 5% y al 95% de probabilidad, y los histogramas de las posiciones y los tiempos de llegada.

En las figuras 21 a 26 mostramos, a la izquierda, tres campos escalados de transmisividad, y a la derecha los correspondientes campos de alturas piezométricas que resultan de la resolución del modelo de flujo. Sobre todos los campos hemos representado gráficamente las trayectorias seguidas por 20 partículas que escapan del repositorio.

En la figura 27 se muestran la curva de llegada media y las envolventes correspondiente al 5% y al 95% de probabilidad para cada uno de los métodos considerados. También se ha representado la curva de llegada correspondiente al campo de referencia. Por último en las figuras 28 y 29 pueden verse los histogramas de la posición y los tiempos de llegada respectivamente, correspondientes al campo de referencia y a cada uno de los métodos.

7.2.1 Análisis de los resultados

Simulación gaussiana secuencial

La simulación de la variable principal adoptando un modelo multigaussiano para describir su dependencia espacial fue realizada considerando sólo la información acerca de la variable principal, esto es, los 20 datos duros. A pesar de esto, y como hemos comentado en el punto 6.1.1, la simulación gaussiana secuencial de una variable puede realizarse considerando una o varias variables secundarias, siendo también posible la simulación de varios atributos. Aquí estamos interesados en evaluar el impacto que tiene la incorporación de la información complementaria comparando los resultados obtenidos en este punto con los de los subsiguientes.

Simulación gaussiana secuencial por cokrigado colocalizado bajo un modelo markoviano

La simulación gaussiana del atributo principal se realiza considerando además de los 20 datos duros como en el método anterior, el dato secundario colocalizado en el punto a estimar. Si comparamos las figuras 21 y 22 podemos ver la notable influencia que tiene considerar información complementaria. Los campos simulados de transmisividad son más consistentes con el de referencia, lo que naturalmente se refleja en los correspondientes campos de alturas piezométricas. Las características a gran escala son mejor descritas cuando se considera la información secundaria por lo que la incertidumbre en la predicción de la posición y de los tiempos de llegada será menor como puede apreciarse observando las curvas de llegada representadas en la figura 27 para cada método. Ambas curvas medias muestran la misma desviación sesgada con respecto a la de referencia, pero el ancho del intervalo de probabilidad definido por las curvas envolventes es mucho menor para el caso que considera información secundaria. Más interesante es observar el comportamiento de la trayectorias de las partículas de contaminante. Mientras que para el caso en que sólo se considera la información principal la posición de llegada puede variar considerablemente de un campo a otro, en el caso de incorporar información secundaria el rango de variación de estas posiciones disminuye drásticamente. Esto puede verse claramente expresado por los histogramas de la posi-

ción de llegada de la figura 28. La incorporación de la información secundaria reduce en más de 100 veces la varianza de la posición de llegada. Algo similar ocurre con los tiempos aunque en menor escala.

Simulación indicadora secuencial

La simulación indicadora secuencial se realizó considerando toda la información disponible. En este caso los datos blandos fueron codificados como datos de intervalo a partir del diagrama de dispersión de los datos de calibración como ha sido descrito en 7.1. En la figura 23 se muestra el resultado de nuestro ejercicio para este método. Este método tampoco logra una adecuada reproducción de la variabilidad a gran escala de la transmisividad, lo que afecta a las trayectorias, y en consecuencia a los tiempos y la posición de llegada. Esto puede comprobarse observando en la figura 27 las curvas de llegada y sobre todo las envolventes para el 5% y el 95% de probabilidad. A pesar de que la curva media se aparta de igual manera que para los demás métodos, la incertidumbre representada por la separación entre envolventes es mayor. De igual forma lo expresa el histograma correspondiente en la figura 28.

Simulación secuencial por cokrigado indicador con un modelo de Markov-Bayes

Como hemos explicado en el punto 5.8 el cokrigado indicador con un modelo de Markov-Bayes permite la mejor incorporación de la información secundaria. Tres de los campos generados bajo este formalismo se muestran en la figura 24. Puede verse que la reproducción del campo de trans-

misividad es muy bueno lo que redundará en el comportamiento de los tiempos de llegada. Las envolventes en la figura 27 determinan un intervalo más estrecho alrededor de la media y el histograma de la posición de llegada en la figura 28 muestra que los estadísticos globales del conjunto de 100 simulaciones son los más cercanos a los de referencia.

Simulación por campos de probabilidades

La simulación por campo de probabilidades fue implementada utilizando como funciones locales de distribución condicionadas a la información disponible, las construidas vía el algoritmo de cokrigado indicador bajo un formalismo de Markov-Bayes. El método se comporta relativamente bien en cuanto a la representación de la curva media de llegadas, pero no así con respecto a las envolventes. Esto se puede visualizar también en el histograma de las posiciones y tiempos de llegada (figura 28 y 29). El atractivo de este método está en el menor tiempo de CPU que requiere.

Simulación por recocido simulado

Una técnica de recocido simulado ha sido utilizada para generar campos de transmisividad condicionados a los datos duros, al patrón de continuidad espacial y al histograma bivariado obtenido a partir de los datos de calibración. Dado que este es un algoritmo de optimización y que contamos con información secundaria exhaustiva los campos simulados son muy similares entre sí, resultando en que las curvas envolventes están muy cerca de su media.

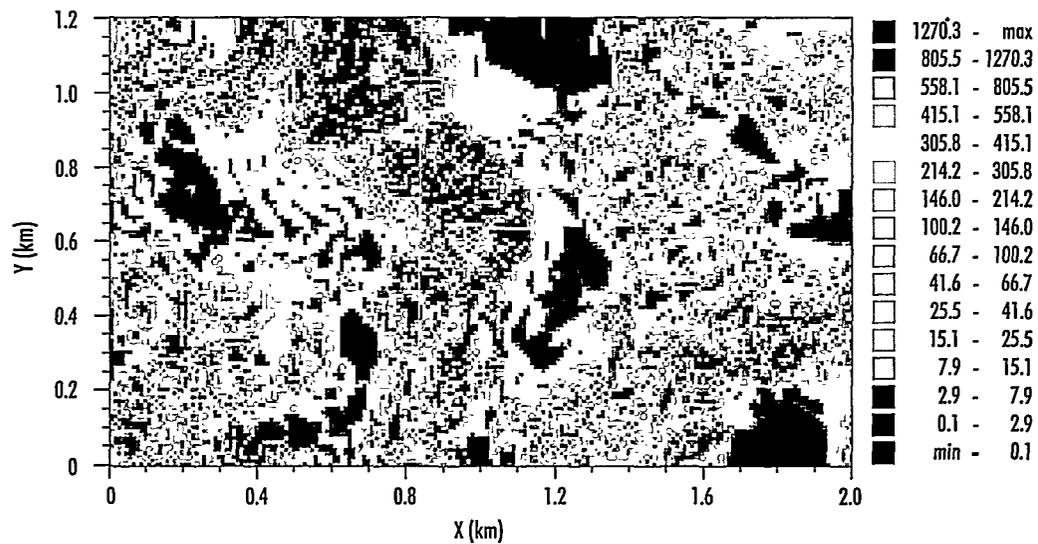


Figura 3. Campo de referencia de la variable principal (U).

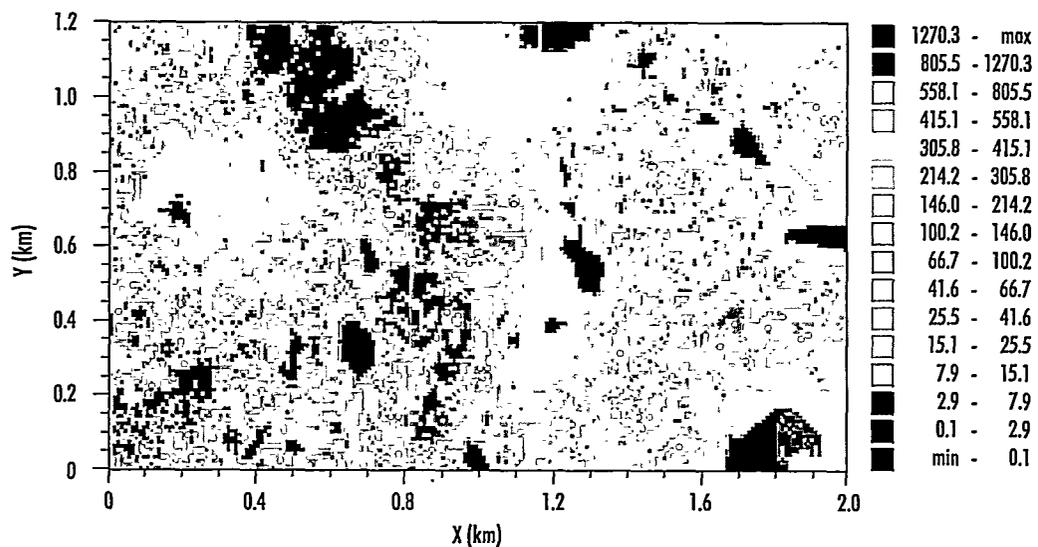


Figura 4. Campo de referencia de la variable secundaria (V).

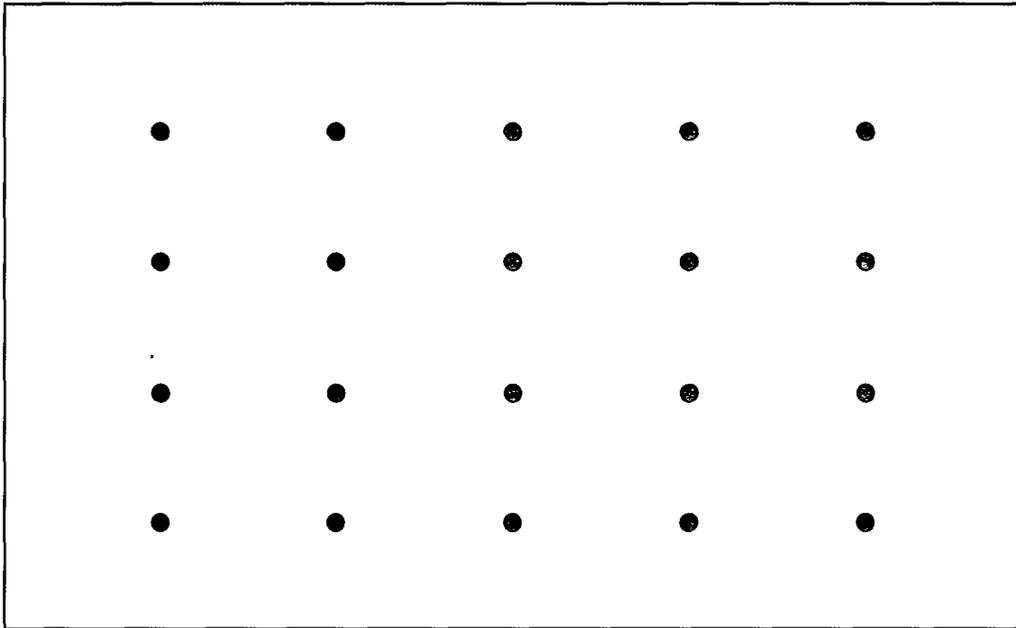


Figura 5. Localización de los datos duros.

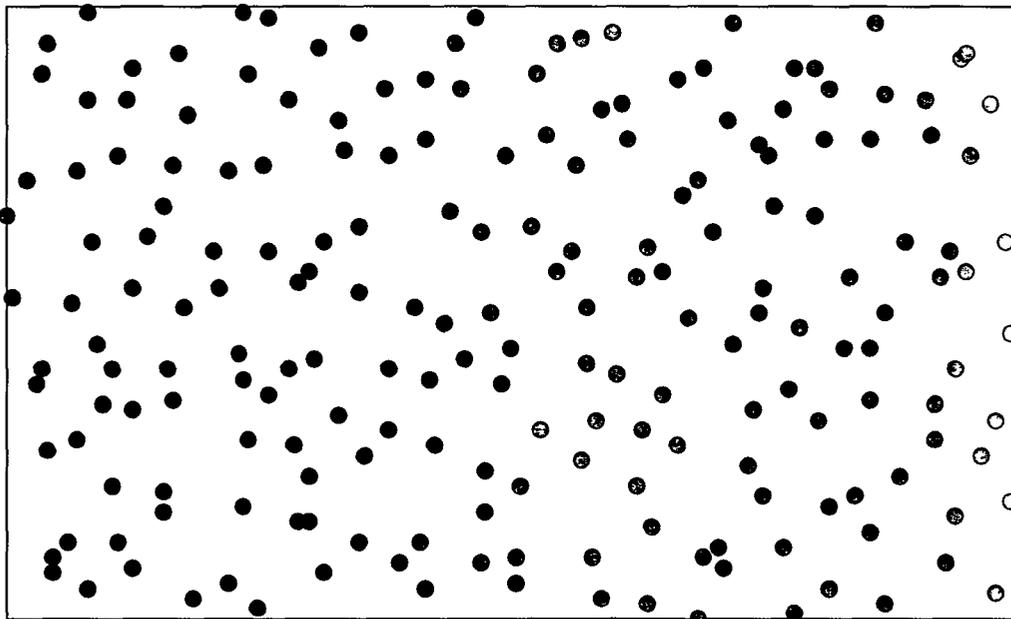


Figura 6. Localización del conjunto de datos de calibración.

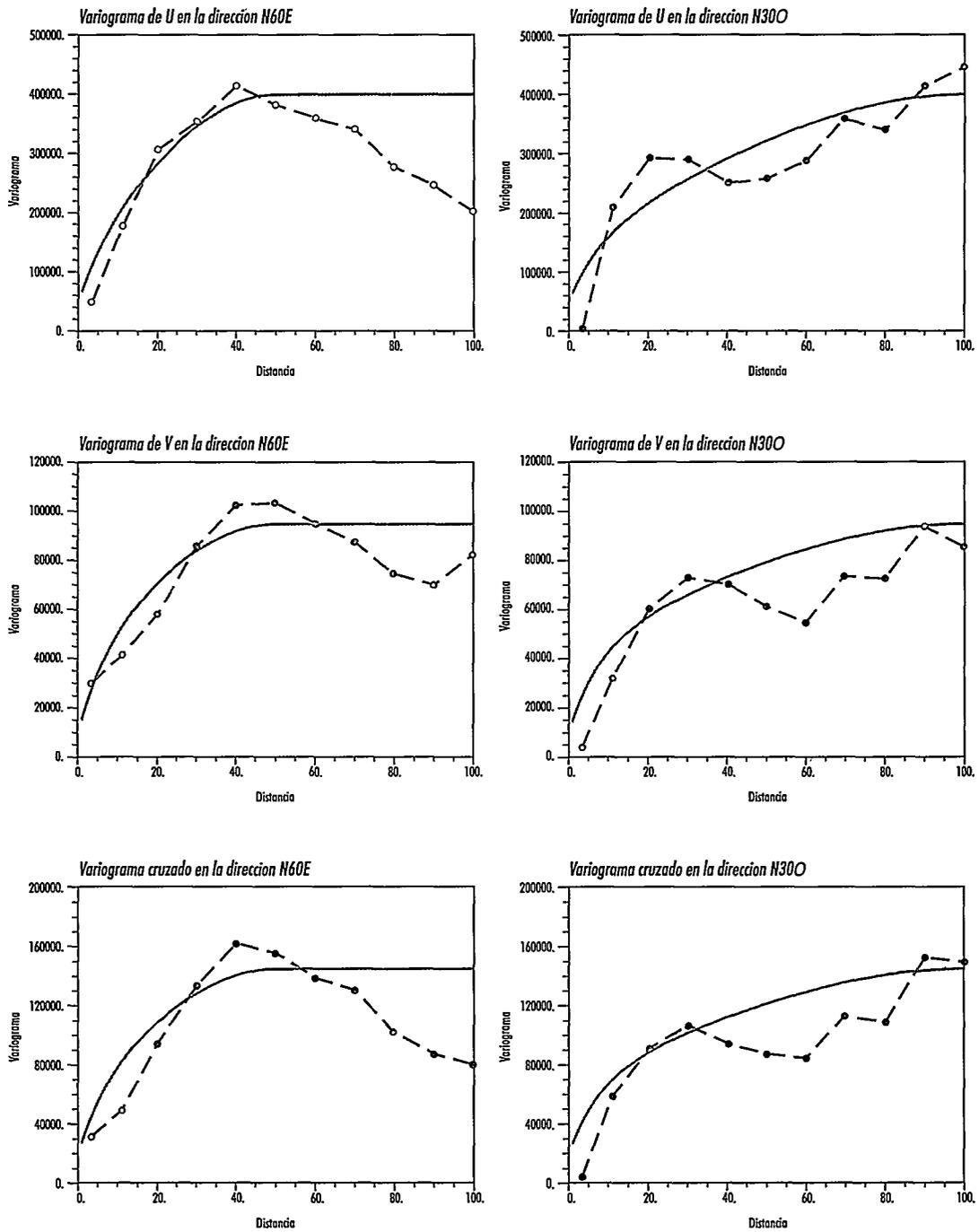


Figura 7. Variogramas experimentales y modelos ajustados. Por columnas los variogramas en las direcciones N60E y N300. Por filas los variogramas de la variable principal, secundaria y cruzada.

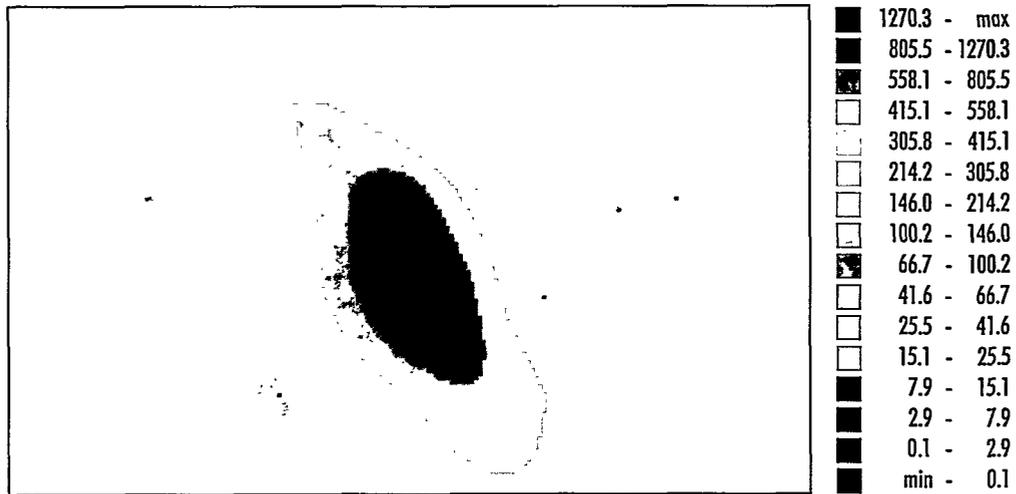


Figura 8. Campo interpolado por kriging simple con media global de los 20 datos duros.

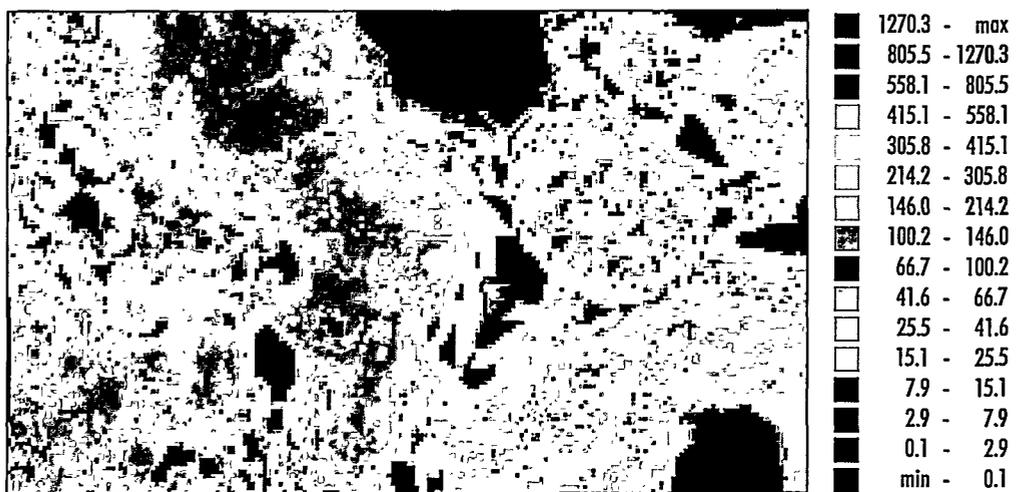


Figura 9. Campo estimado por regresión lineal.

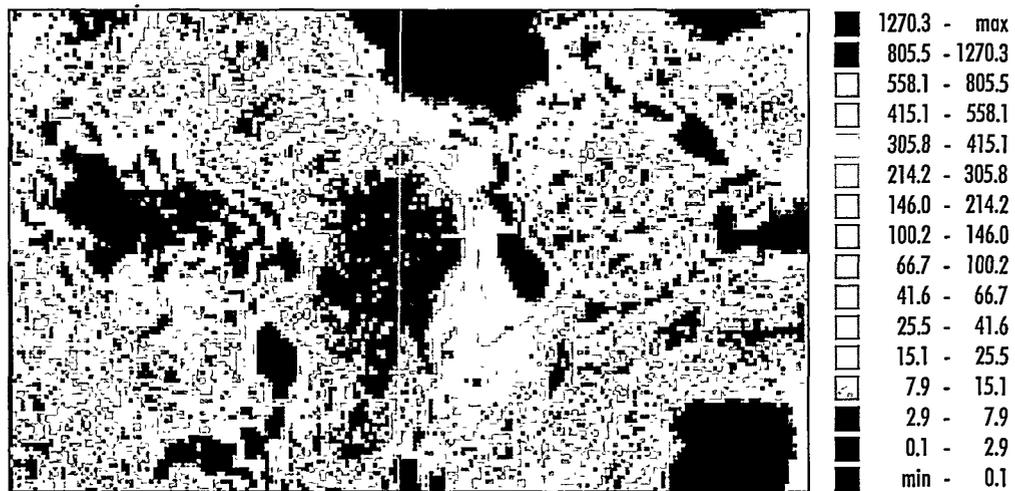


Figura 10. Campo estimado por krigeado con una deriva externa.

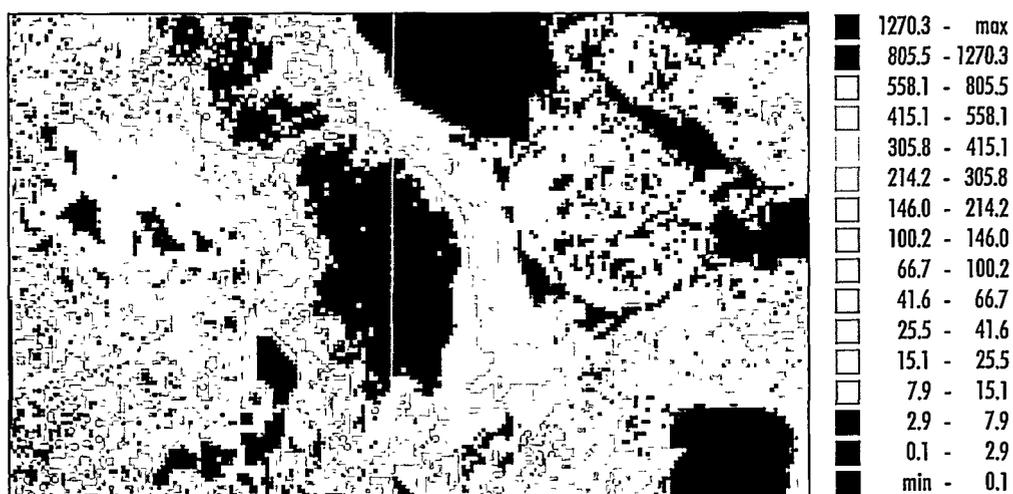


Figura 11. Campo estimado por cokrigeado colocalizado.

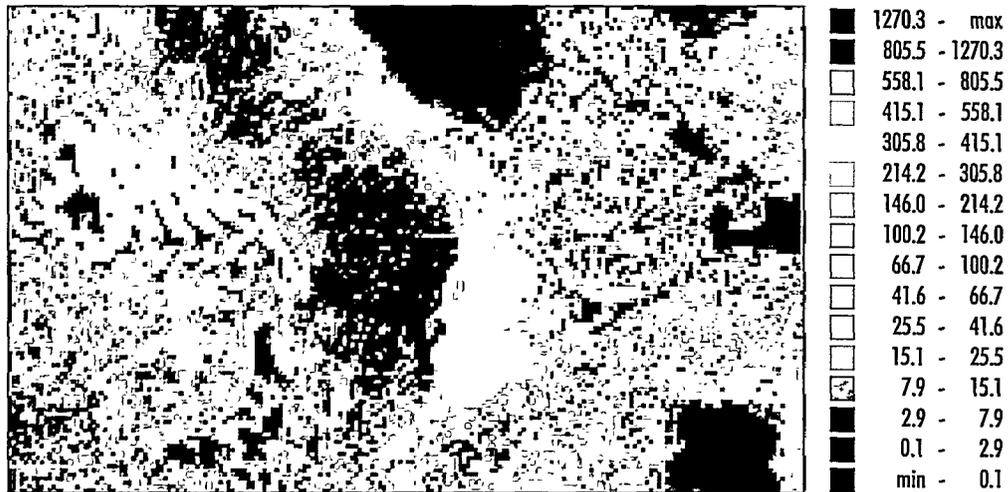


Figura 12. Campo estimado por cokrigado completo.

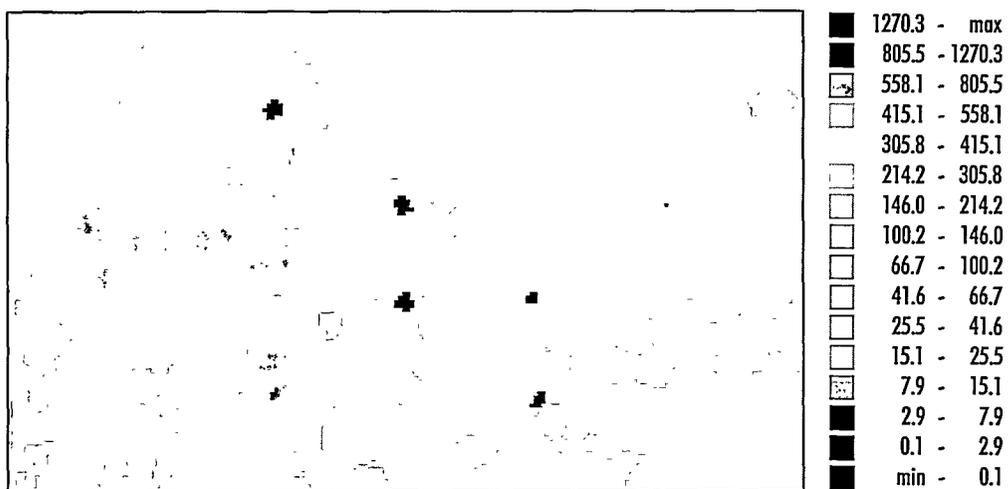


Figura 13. Campo estimado por krigeado indicador.

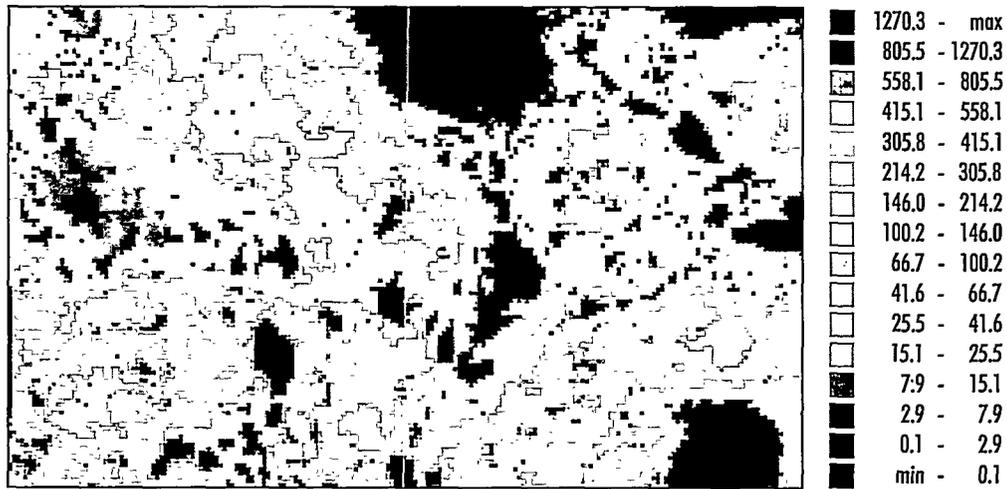


Figura 14. Campo estimado por cokrigado indicador.

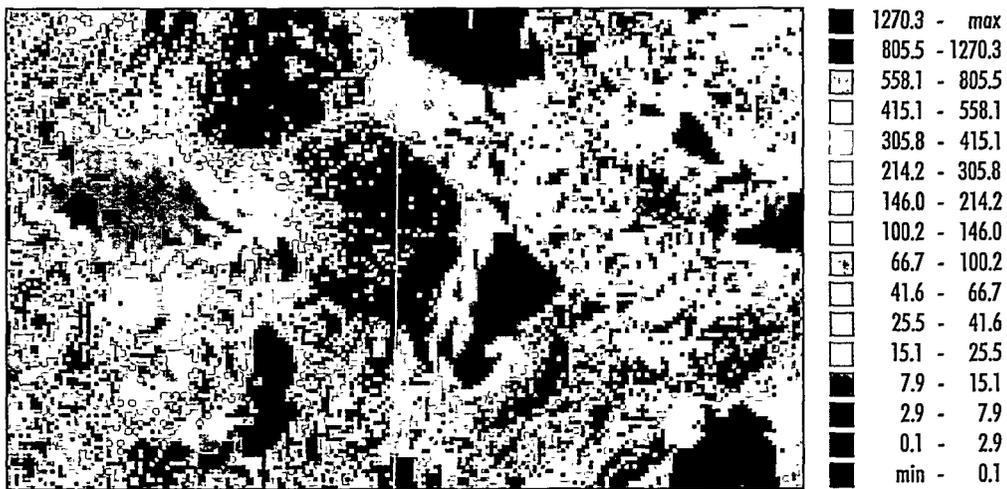


Figura 15. Campo estimado por campo de probabilidades.

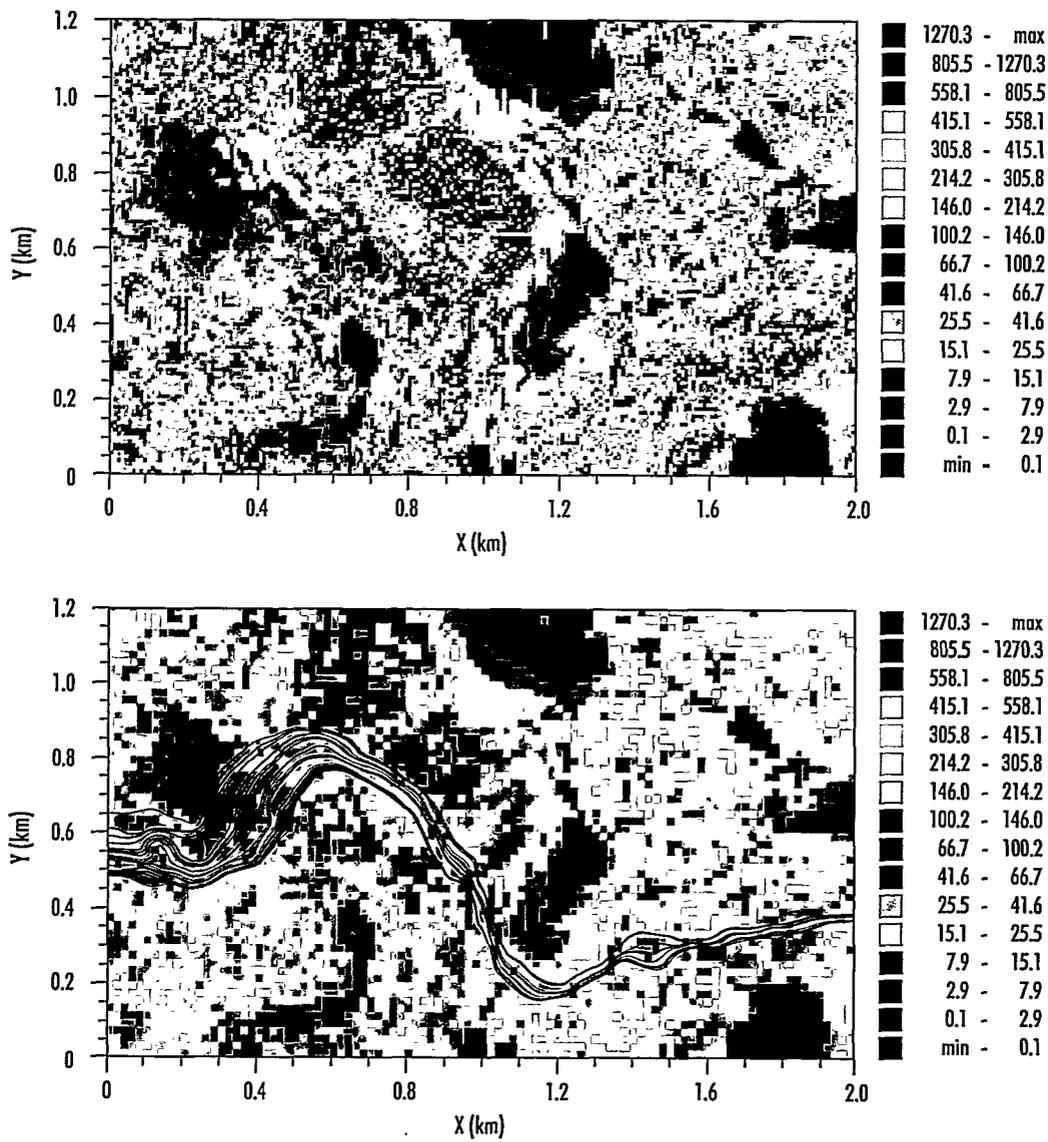


Figura 16. Campo de transmisividades de referencia. Arriba: campo original. Abajo: campo escalado considerando bloques de 2×2 celdas. Sobre este último se han representado gráficamente las trayectorias de 20 partículas.

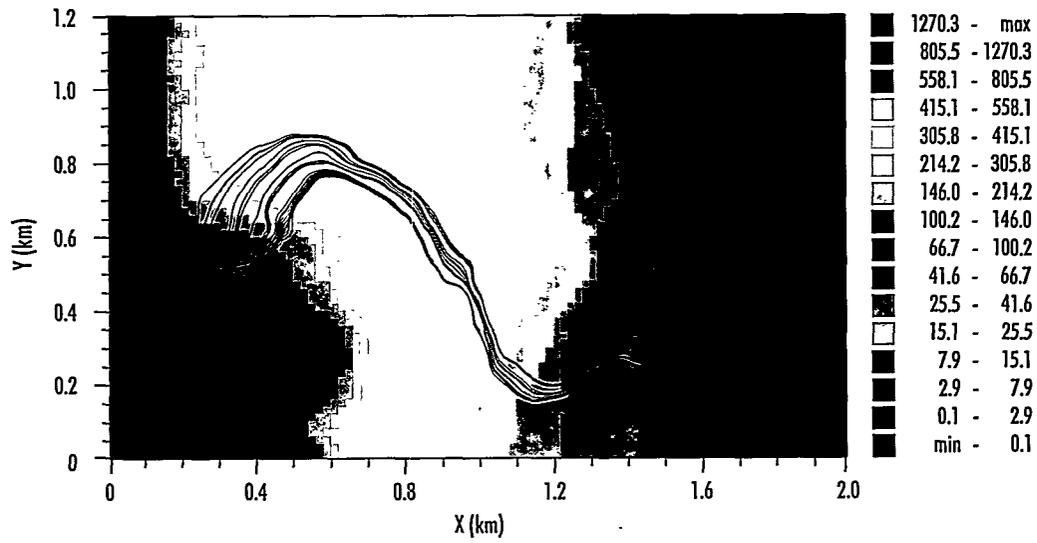


Figura 17. Campo de alturas piezométricas de referencia y trayectorias de 20 partículas.

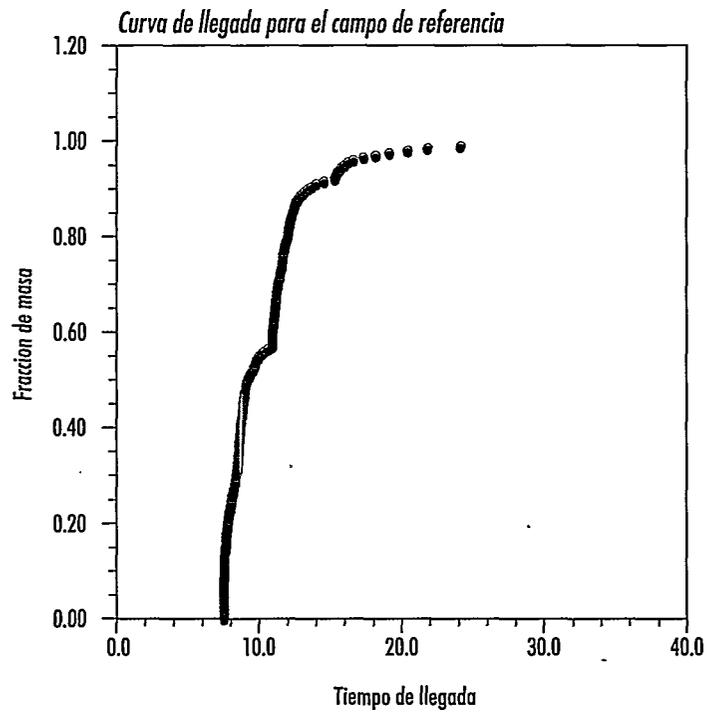


Figura 18. Curva de llegada para el campo de referencia.

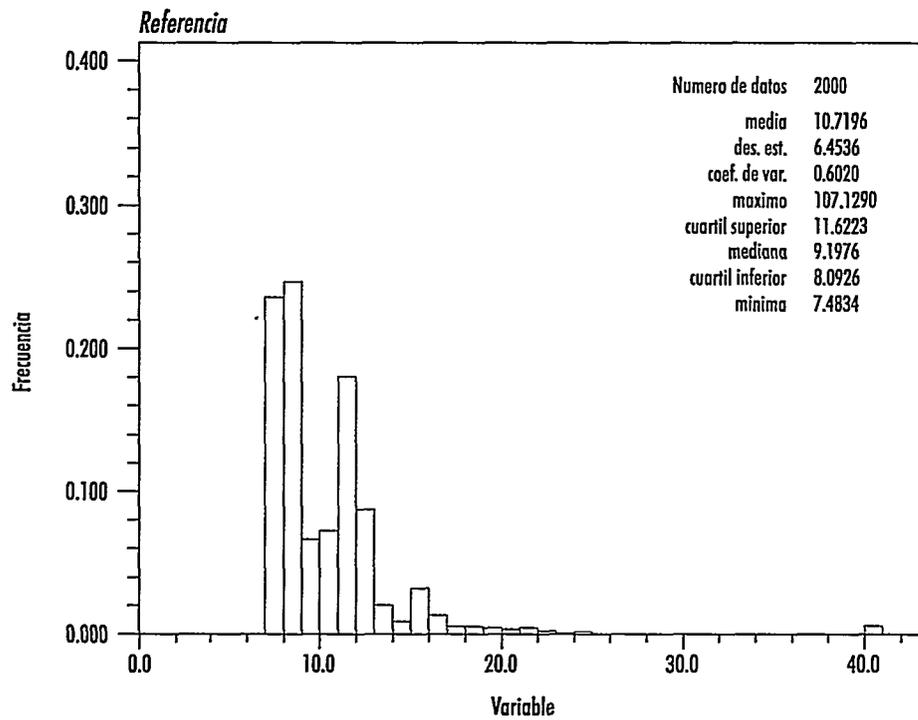


Figura 19. Histograma de los tiempos de llegada para el campo de referencia.

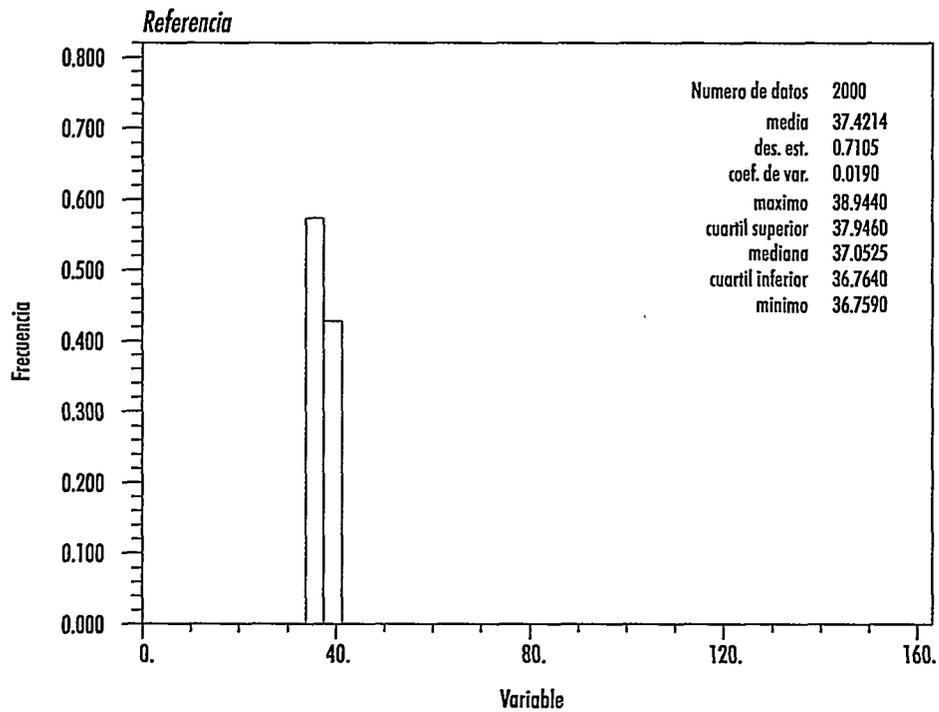


Figura 20. Histograma de la posición de llegada para el campo de referencia.

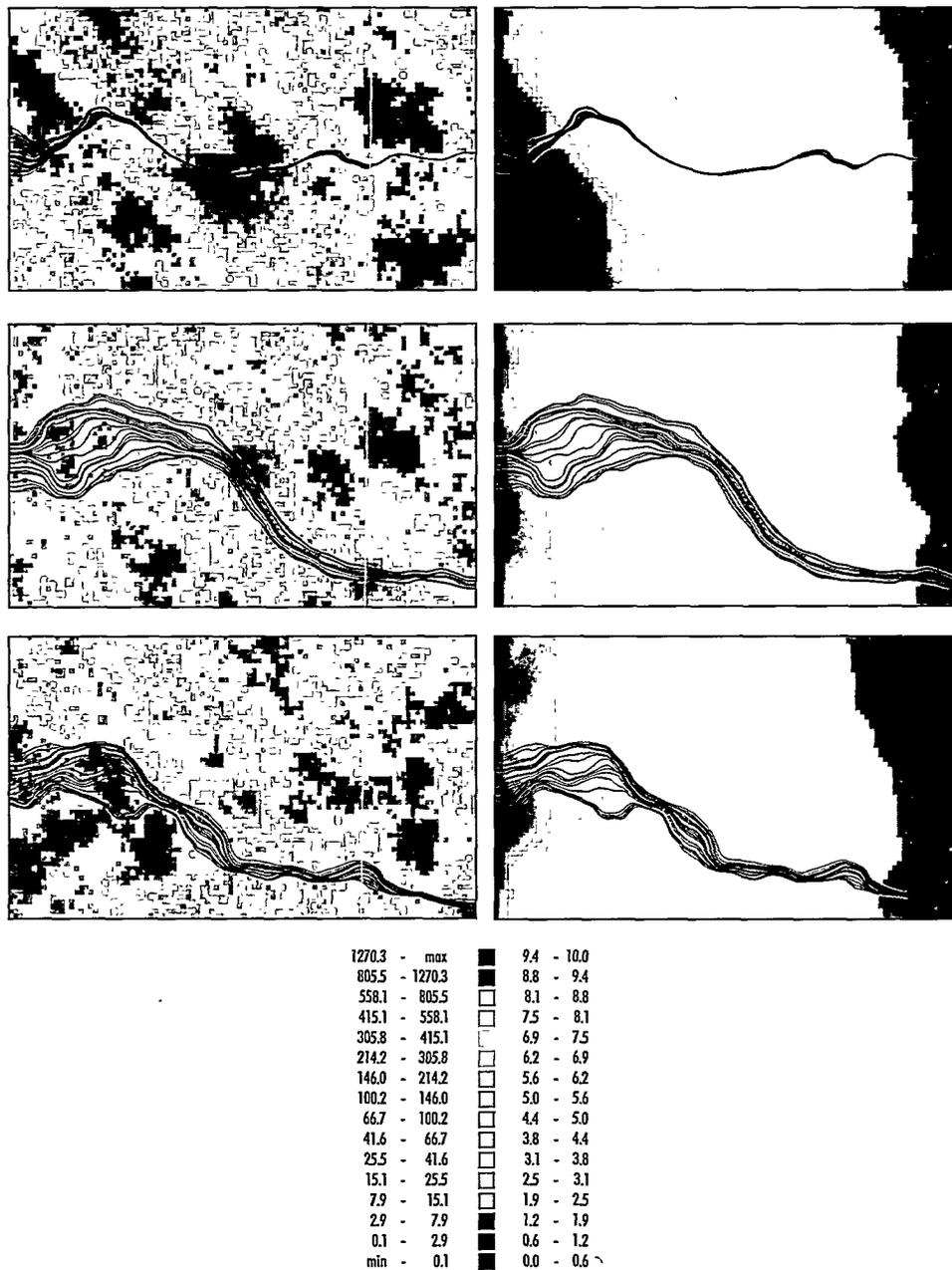


Figura 21. A la izquierda: tres campos generados por simulación gaussiana secuencial. A la derecha: campos de alturas piezométricas resultantes. Sobre todos los campos se han representado las trayectorias para 20 partículas.

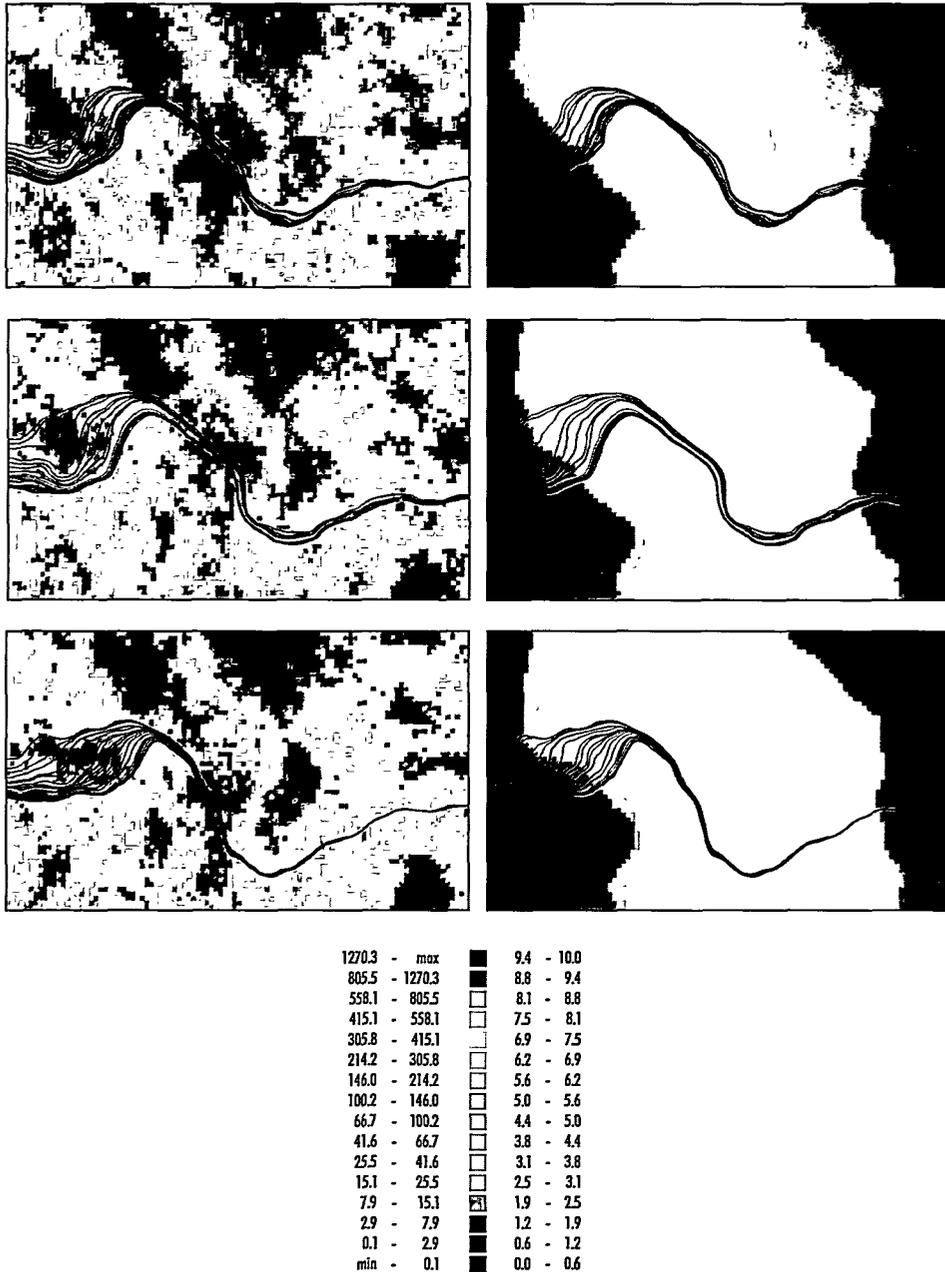


Figura 22. A la izquierda: tres campos generados por simulación gaussiana secuencial por cokrigado colocalizado bajo un modelo markoviano. A la derecha: campos de alturas piezométricas resultantes. Sobre todos los campos se han representado las trayectorias para 20 partículas.

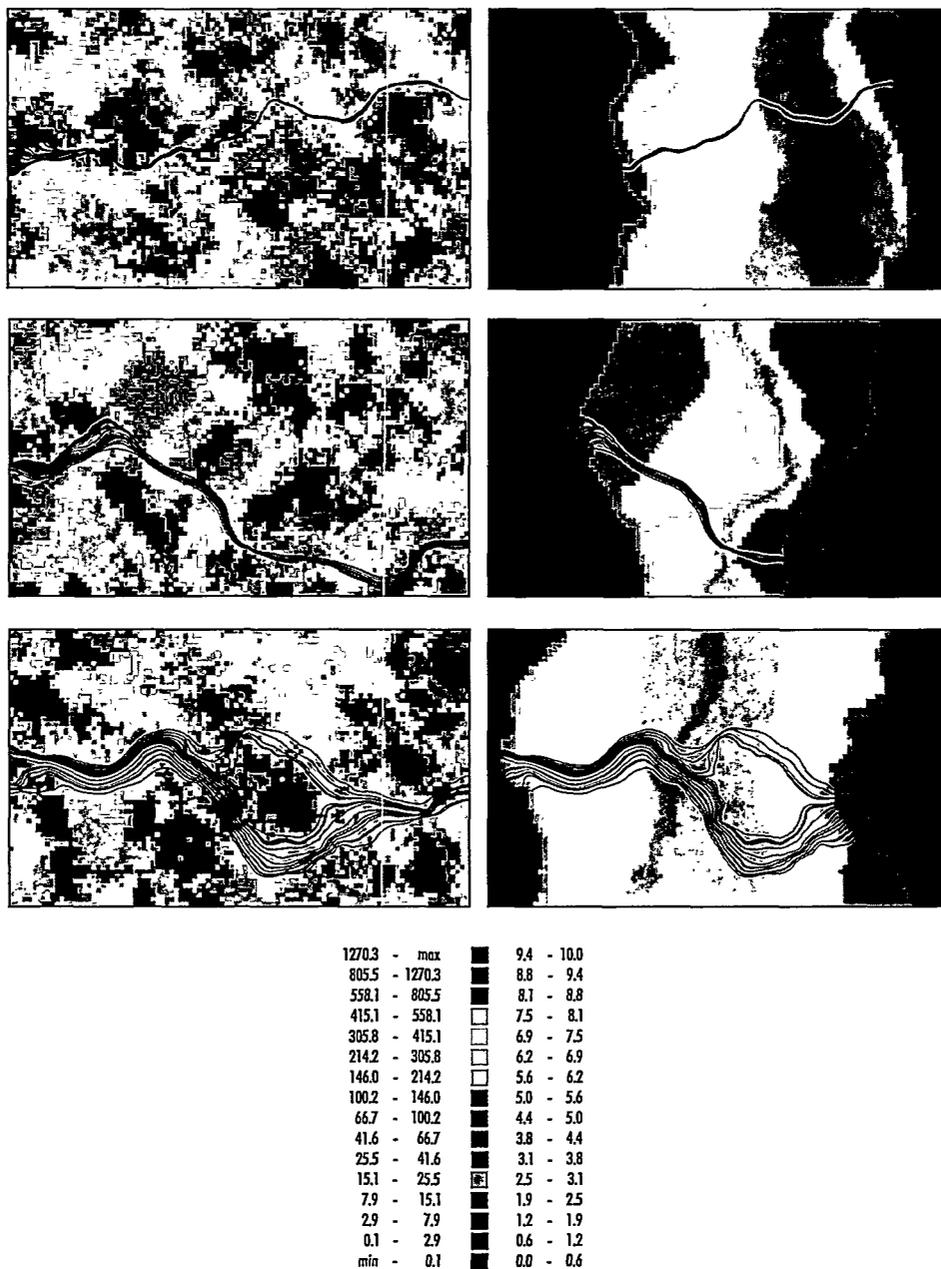


Figura 23. A la izquierda: tres campos generados por simulación indicadora secuencial. A la derecha: campos de alturas piezométricas resultantes. Sobre todos los campos se han representado las trayectorias para 20 partículas.

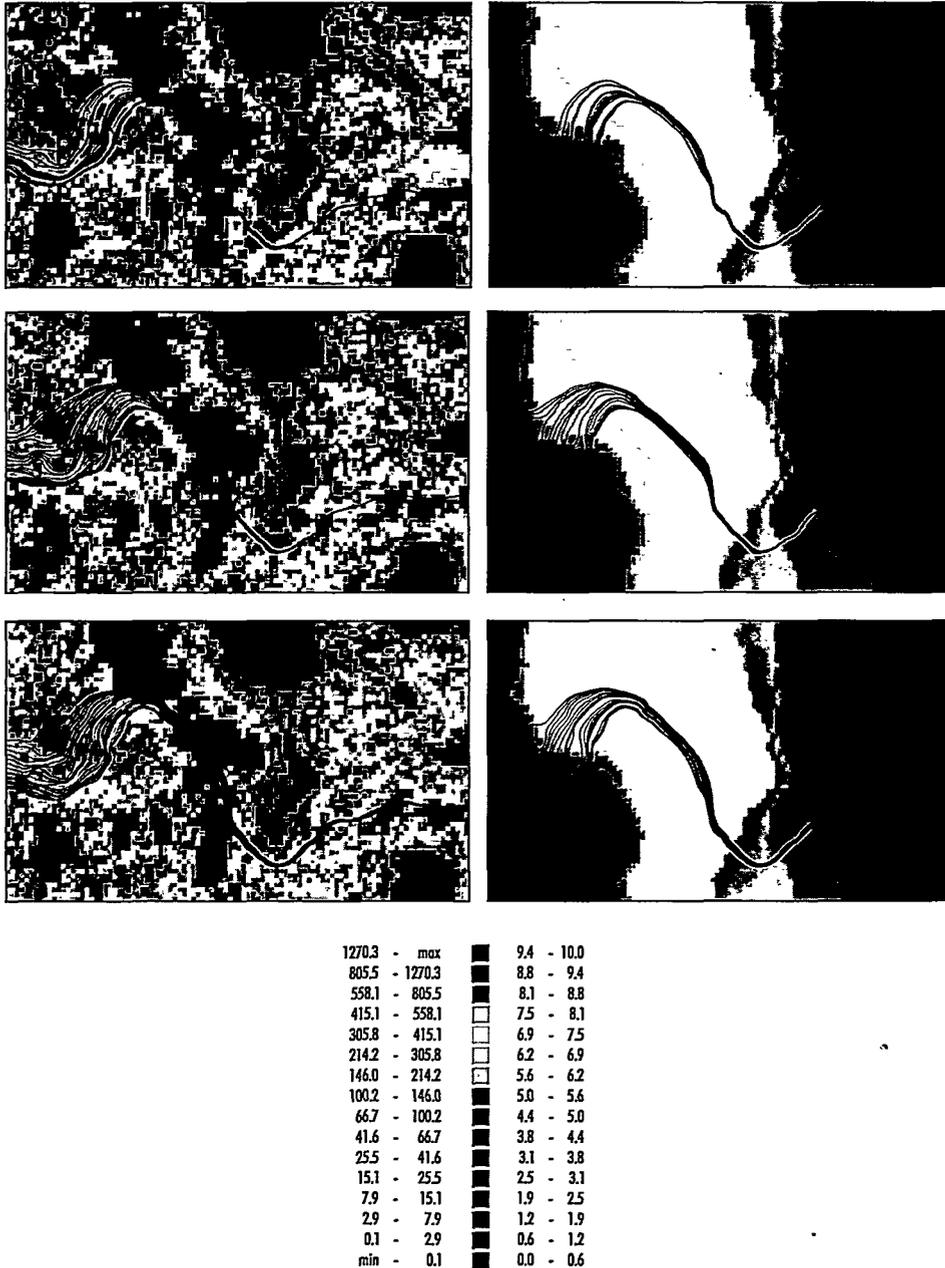
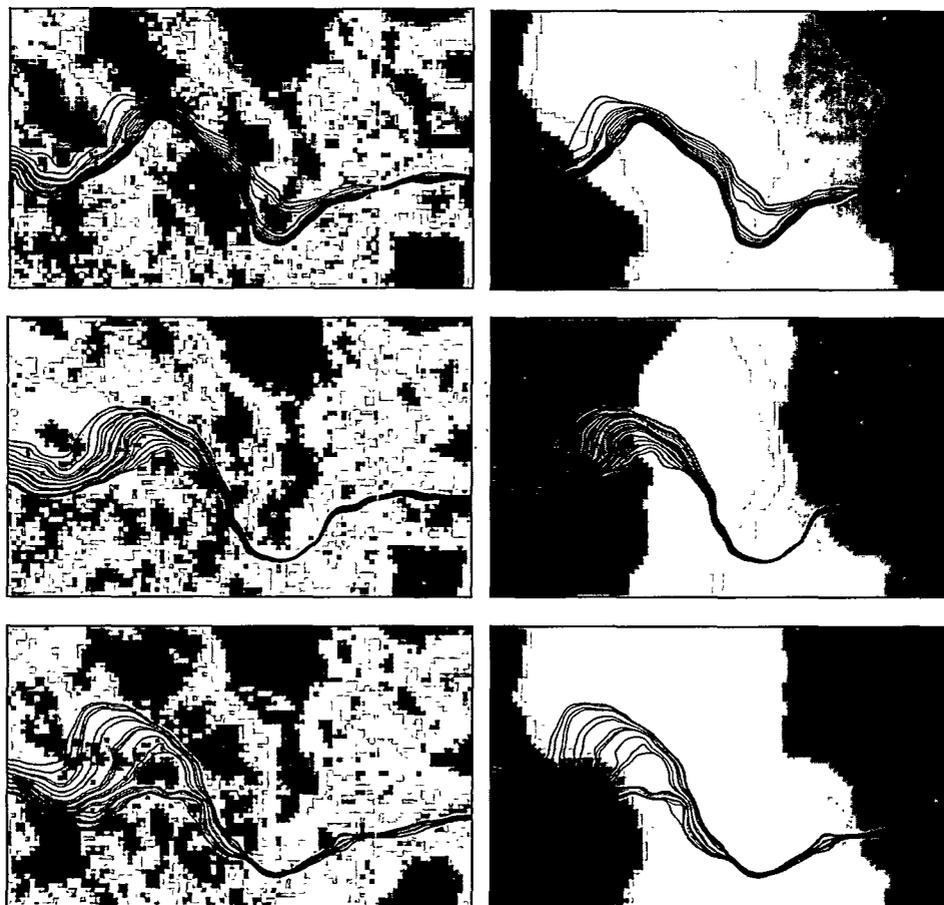
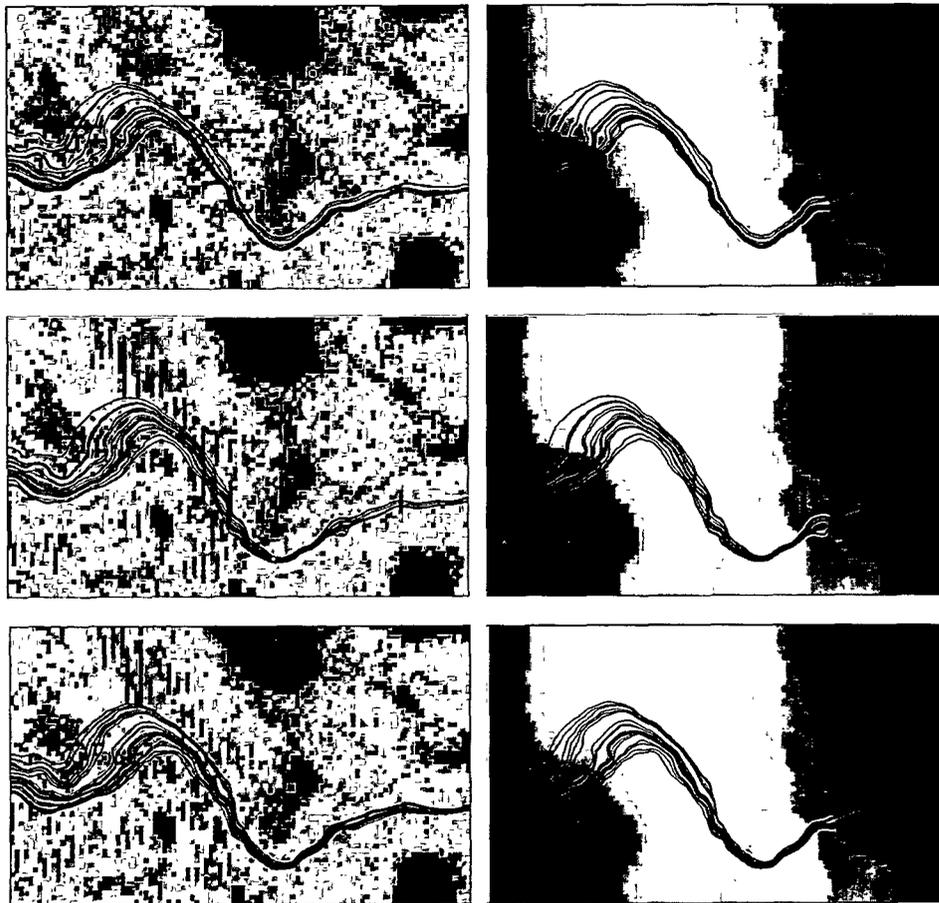


Figura 24. A la izquierda: tres campos generados por simulación secuencial por cokrigado indicador bajo un modelo markoviano. A la derecha: campos de alturas piezométricas resultantes. Sobre todos los campos se han representado las trayectorias para 20 partículas.



1270.3 - max	■	9.4 - 10.0
805.5 - 1270.3	■	8.8 - 9.4
558.1 - 805.5	■	8.1 - 8.8
415.1 - 558.1	■	7.5 - 8.1
305.8 - 415.1	■	6.9 - 7.5
214.2 - 305.8	■	6.2 - 6.9
146.0 - 214.2	■	5.6 - 6.2
100.2 - 146.0	■	5.0 - 5.6
66.7 - 100.2	■	4.4 - 5.0
41.6 - 66.7	■	3.8 - 4.4
25.5 - 41.6	■	3.1 - 3.8
15.1 - 25.5	■	2.5 - 3.1
7.9 - 15.1	■	1.9 - 2.5
2.9 - 7.9	■	1.2 - 1.9
0.1 - 2.9	■	0.6 - 1.2
min - 0.1	■	0.0 - 0.6

Figura 25. A la izquierda: tres campos generados por simulación por campo de probabilidades. A la derecha: campos de alturas piezométricas resultantes. Sobre todos los campos se han representado las trayectorias para 20 partículas.



1270.3 - max	■	9.4 - 10.0
805.5 - 1270.3	■	8.8 - 9.4
558.1 - 805.5	■	8.1 - 8.8
415.1 - 558.1	■	7.5 - 8.1
305.8 - 415.1	■	6.9 - 7.5
214.2 - 305.8	■	6.2 - 6.9
146.0 - 214.2	■	5.6 - 6.2
100.2 - 146.0	■	5.0 - 5.6
66.7 - 100.2	■	4.4 - 5.0
41.6 - 66.7	■	3.8 - 4.4
25.5 - 41.6	■	3.1 - 3.8
15.1 - 25.5	■	2.5 - 3.1
7.9 - 15.1	■	1.9 - 2.5
2.9 - 7.9	■	1.2 - 1.9
0.1 - 2.9	■	0.6 - 1.2
min - 0.1	■	0.0 - 0.6

Figura 26. A la izquierda: tres campos generados por simulación por recocido simulado. A la derecha: campos de alturas piezométricas resultantes. Sobre todos los campos se han representado las trayectorias para 20 partículas.

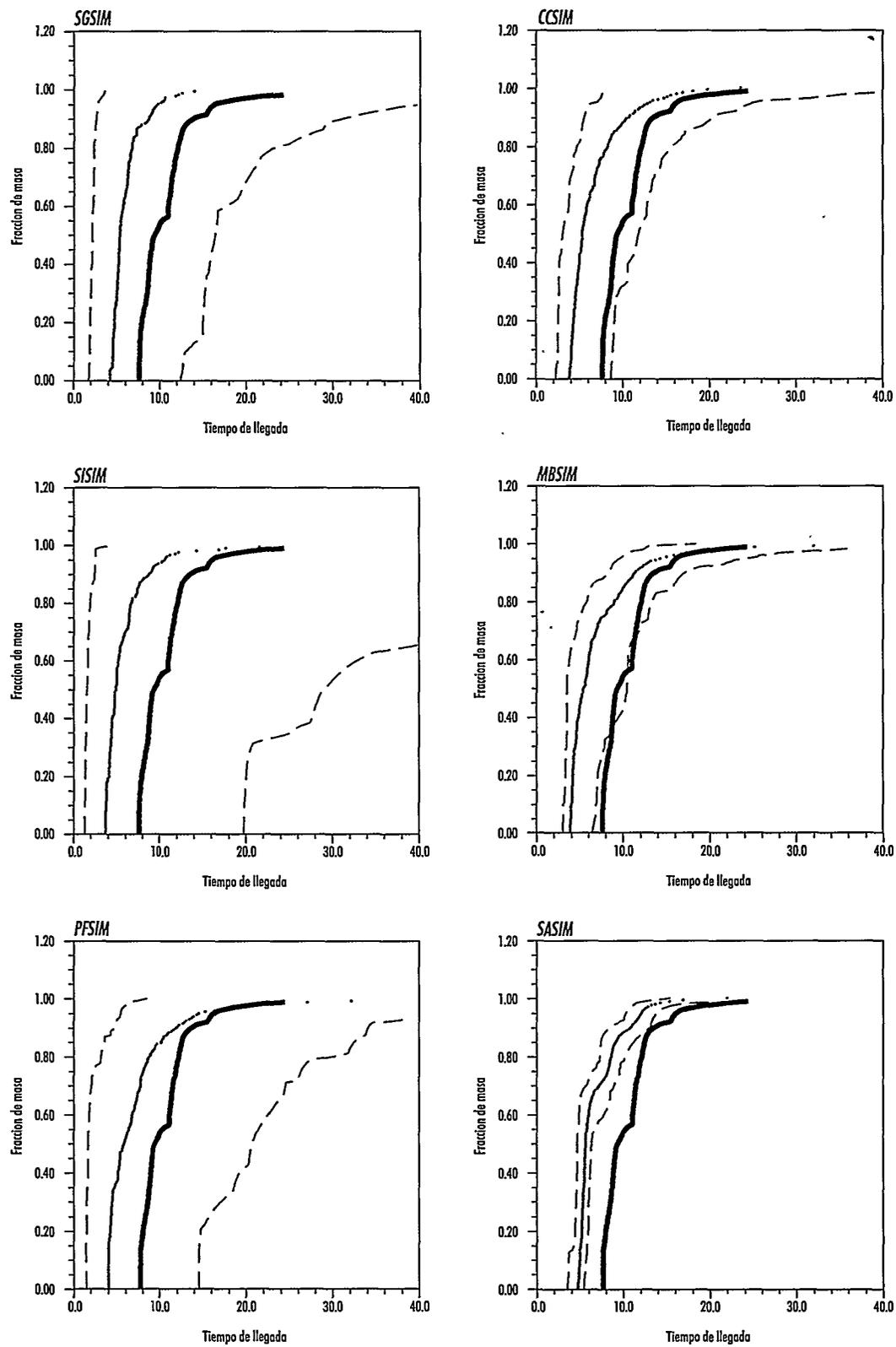


Figura 27. Curvas de llegada para las 100 realizaciones generadas por cada método. En línea continua la curva correspondiente al campo de referencia, en línea punteada la curva media del ensamble, y en línea de trazos los límites para el 5 y el 95% de probabilidad.

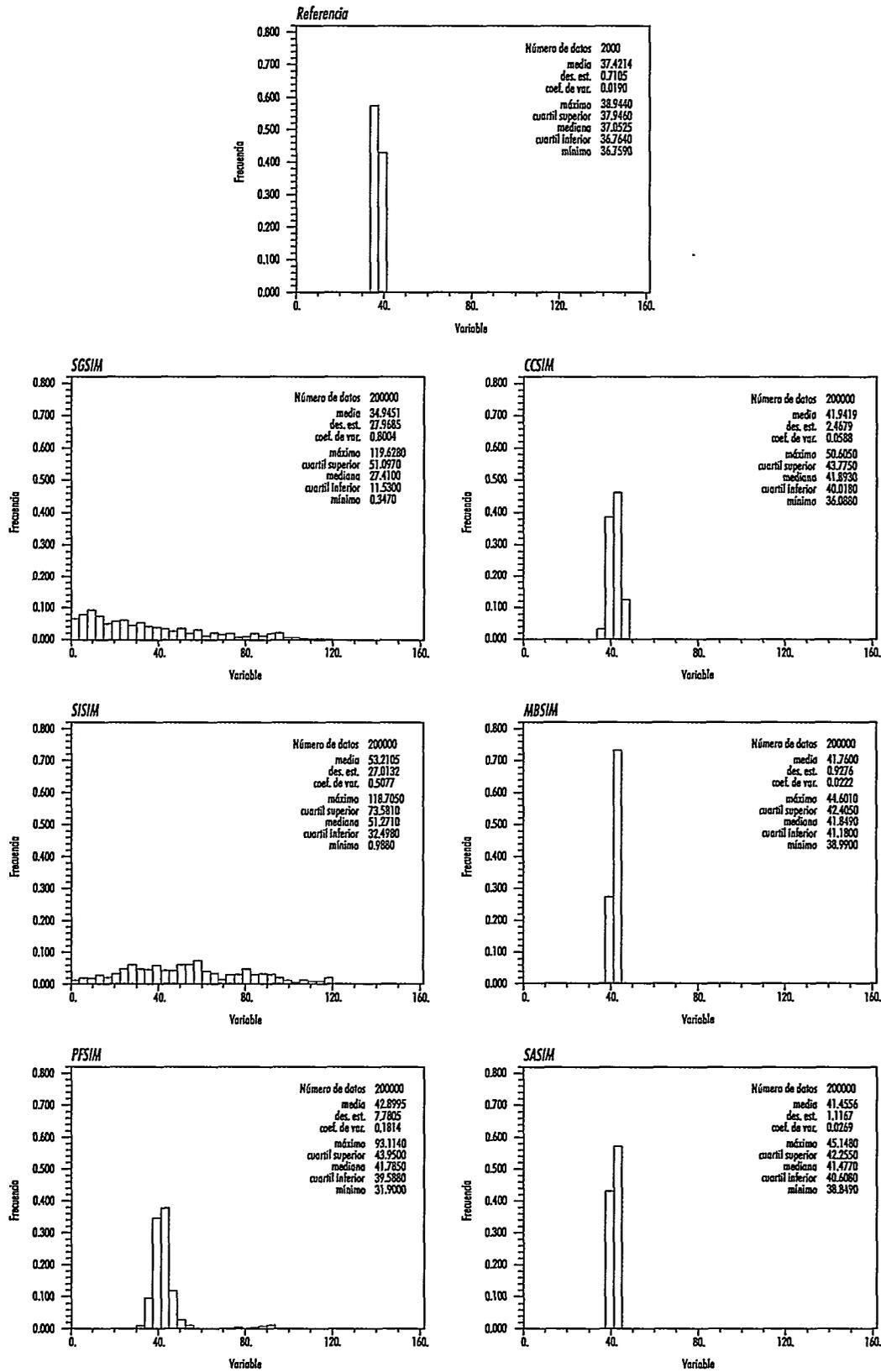


Figura 28. Histogramas de la posición de llegada para el campo de referencia y para las 100 realizaciones generadas por cada método.

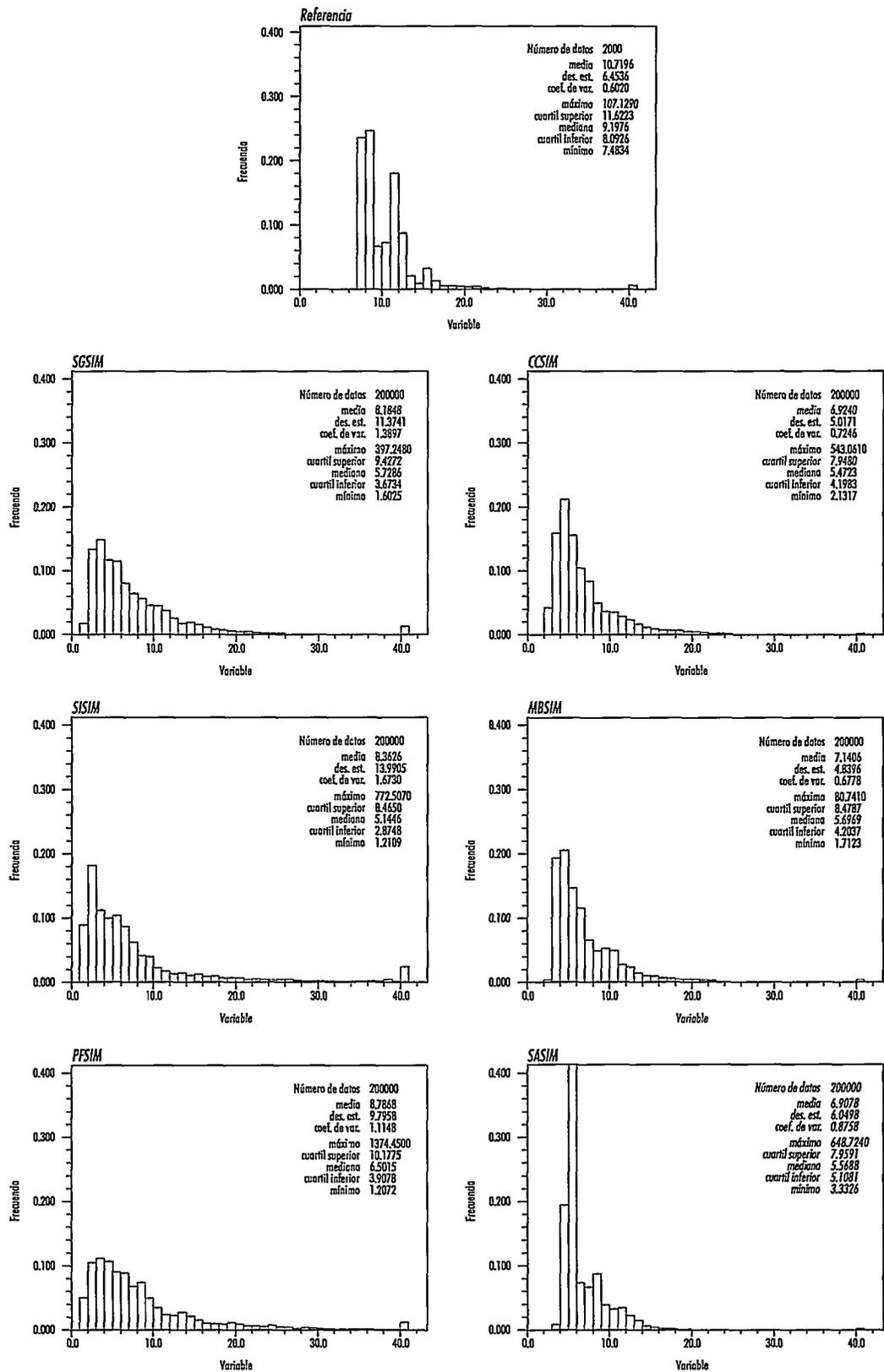
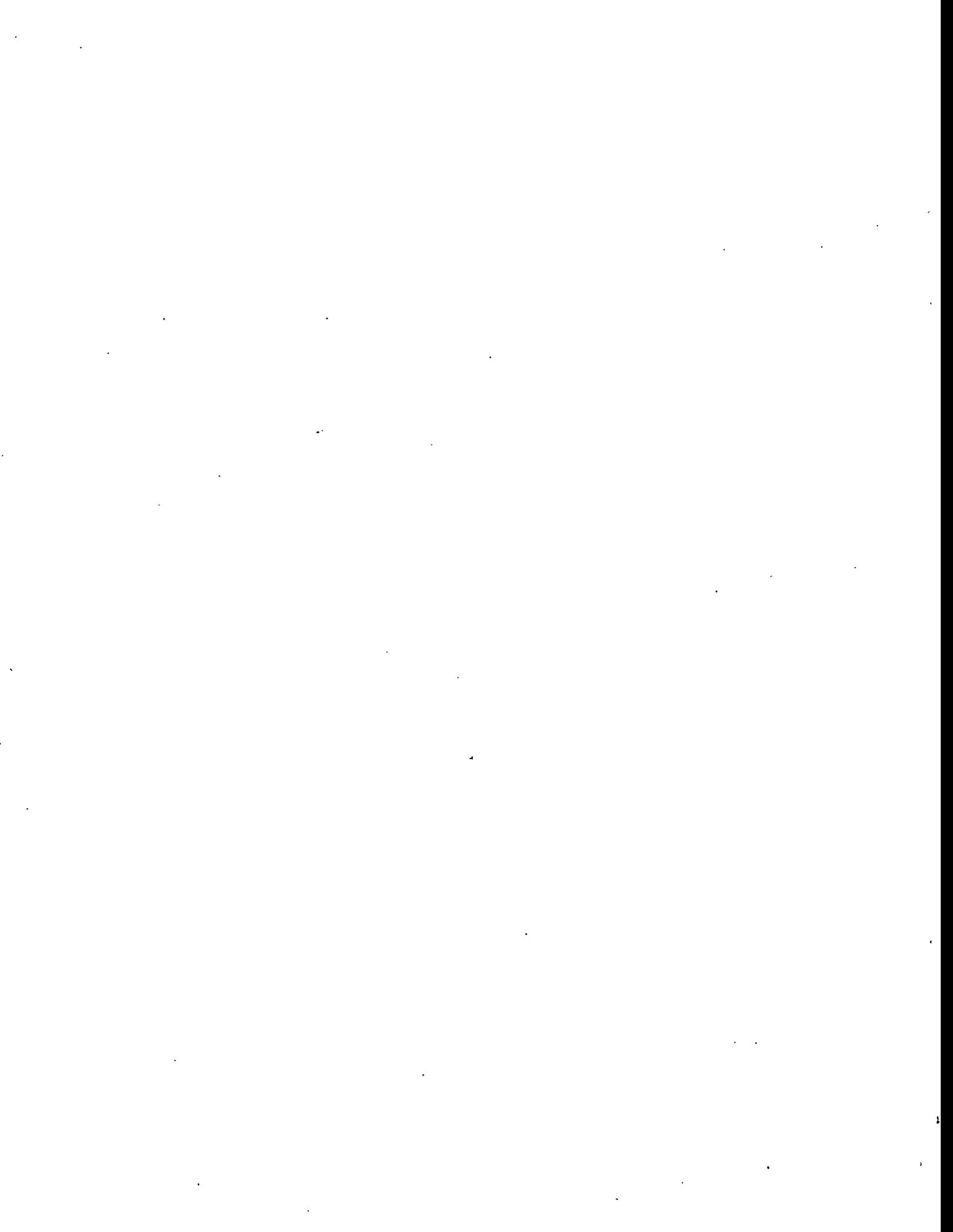


Figura 29. Histogramas de los tiempos de llegada para el campo de referencia y para las 100 realizaciones generadas por cada método.

8

CONCLUSIONES



En este informe hemos revisado con cierto detalle los algoritmos geoestadísticos existentes capaces de incorporar distintos tipos de información. Las técnicas disponibles pueden clasificarse según persigan la obtención de una o múltiples representaciones del parámetro modelizado. Hablamos así de algoritmos de estimación y de algoritmos de simulación respectivamente.

De los algoritmos de estimación, salvo los casos de krigeado simple, ordinario y universal, todos permiten la consideración de varias variables secundarias. La regresión tradicional es el más sencillo de todos estos algoritmos, pero tiene grandes limitaciones para incorporar la información proveniente de una fuente secundaria. Se comporta adecuadamente cuando se dispone de información secundaria exhaustivamente muestreada y bien correlacionada con la principal. El krigeado con una deriva externa permite la inclusión en el campo estimado de una tendencia extraída de una variable secundaria. Sin embargo los valores de esta variable externa no intervienen en la estimación del atributo principal. Además, al utilizar esta técnica, estamos asumiendo que el patrón de variabilidad de la variable secundaria es representativo de la principal, lo que es generalmente difícil de documentar.

El cokrigeado, en cualquiera de sus formas, proporciona un estimador donde se considera de forma separada cada tipo de información. Por un lado la variable secundaria interviene directamente en la inferencia de la primaria, y por otro es posible considerar las autocorrelaciones de cada variable y las correlaciones cruzadas entre ellas. En su implementación tradicional el costo computacional del cokrigeado es elevado. Hay variantes dentro del cokrigeado que permiten reducir la tarea de estimar las covarianzas requeridas por cada sistema de ecuaciones planteado. Mientras que el cokrigeado completo considera, dentro de un cierto entorno, todos los datos disponibles para realizar cada estimación, el cokrigeado colocalizado sólo incorpora como información secundaria el dato localizado en el punto a estimar, reduciendo considerablemente el número de covarianzas a inferir. Si además se asume un modelo de tipo markoviano, estas inferencias se reducen a las covarianzas de la variable principal, ya que las correspondientes a la variable secundaria y las cruzadas se expresan como una función sencilla de las primeras.

La introducción de la función indicadora permite codificar, de manera sencilla y general, diferentes

tipos de datos. Por medio del krigeado indicador se puede inferir, ya no un único valor local, sino un modelo de incertidumbre completo en cada punto considerado, esto es, es posible establecer la distribución local acumulada de probabilidades del parámetro modelizado. A partir de estas distribuciones se puede seleccionar la media o la mediana como valor estimado.

El algoritmo de Markov-Bayes, aplicado a variables indicadoras, considera un estimador cokrigeado para la incorporación de la información externa. Por medio de una asunción markoviana similar a la comentada antes, es posible reducir el número de covarianzas a calcular, de tal forma que en la práctica no necesitamos calcular más covarianzas que en el krigeado indicador estándar. En contrapartida, la consideración de un estimador por cokrigeado tiene las ventajas de poder separar los patrones de continuidad espacial de los distintos tipos de información, y la de permitir la mejora de las funciones de probabilidad acumuladas locales estimadas a priori en función de la información vecina.

De la aplicación de estos algoritmos para la construcción de un sólo campo de la variable de interés podemos concluir que, para el caso en que, como en nuestro ejercicio disponemos de información secundaria densamente muestreada, los algoritmos más sencillos, como la regresión lineal y el krigeado con una deriva externa pueden tener un buen comportamiento a un menor costo computacional que los más complejos.

Cualquiera de los algoritmos de interpolación que permiten la inferencia de un modelo de incertidumbre, puede ser utilizado para simulación estocástica. Se trata sólo de obtener varias realizaciones a partir de las funciones de probabilidad construidas en cada localización. La simulación secuencial se muestra como uno de los más flexibles y potentes algoritmos de simulación. La asunción de un modelo multigaussiano representando la distribución del atributo modelizado, conduce al algoritmo de simulación gaussiana secuencial. Como hemos mostrado es el más simple de los algoritmos dada la sencillez del modelo asumido para las variables en juego. Si consideramos un estimador cokrigeado colocalizado y un modelo de Markov el algoritmo resulta aún más sencillo. La simulación gaussiana individual puede extenderse fácilmente a varias variables, lo que la hace especialmente apta para el problema de la incorporación de varios atributos secundarios.

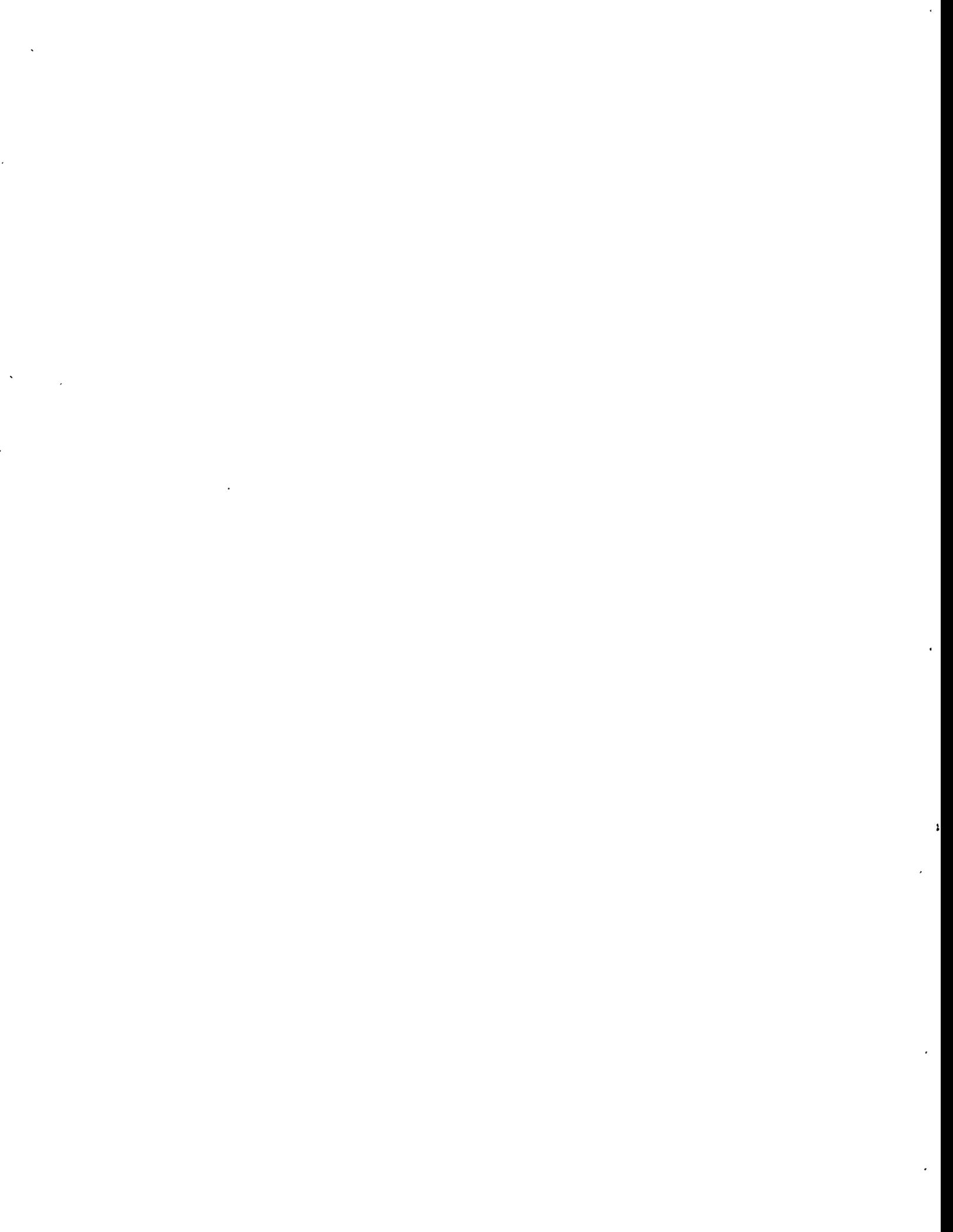
En el ámbito no paramétrico, tanto el krigado indicador como el cokrigado colocalizado bajo un modelo de Markov-Bayes pueden utilizarse para simulación. La simulación por recocido simulado es otra de las técnicas capaces de incorporar información secundaria. La condición para ello es que tal información pueda expresarse como una componente de la función objetivo a ser optimizada. La simulación por campos de probabilidad se para las tareas de construcción de los modelos de incertidumbre y de generación de las realizaciones equiprobables. Esta técnica puede ser atractiva si

contamos con la suficiente información para poder definir dichos modelos en todos los puntos del dominio de simulación.

En cuanto al ejercicio de simulación aquí presentado hemos comprobado que para el caso en que dispongamos de información secundaria exhaustiva, tanto el algoritmo de cokrigado indicador con un modelo de Markov-Bayes como el de simulación por recocido simulado se muestran como los más robustos.

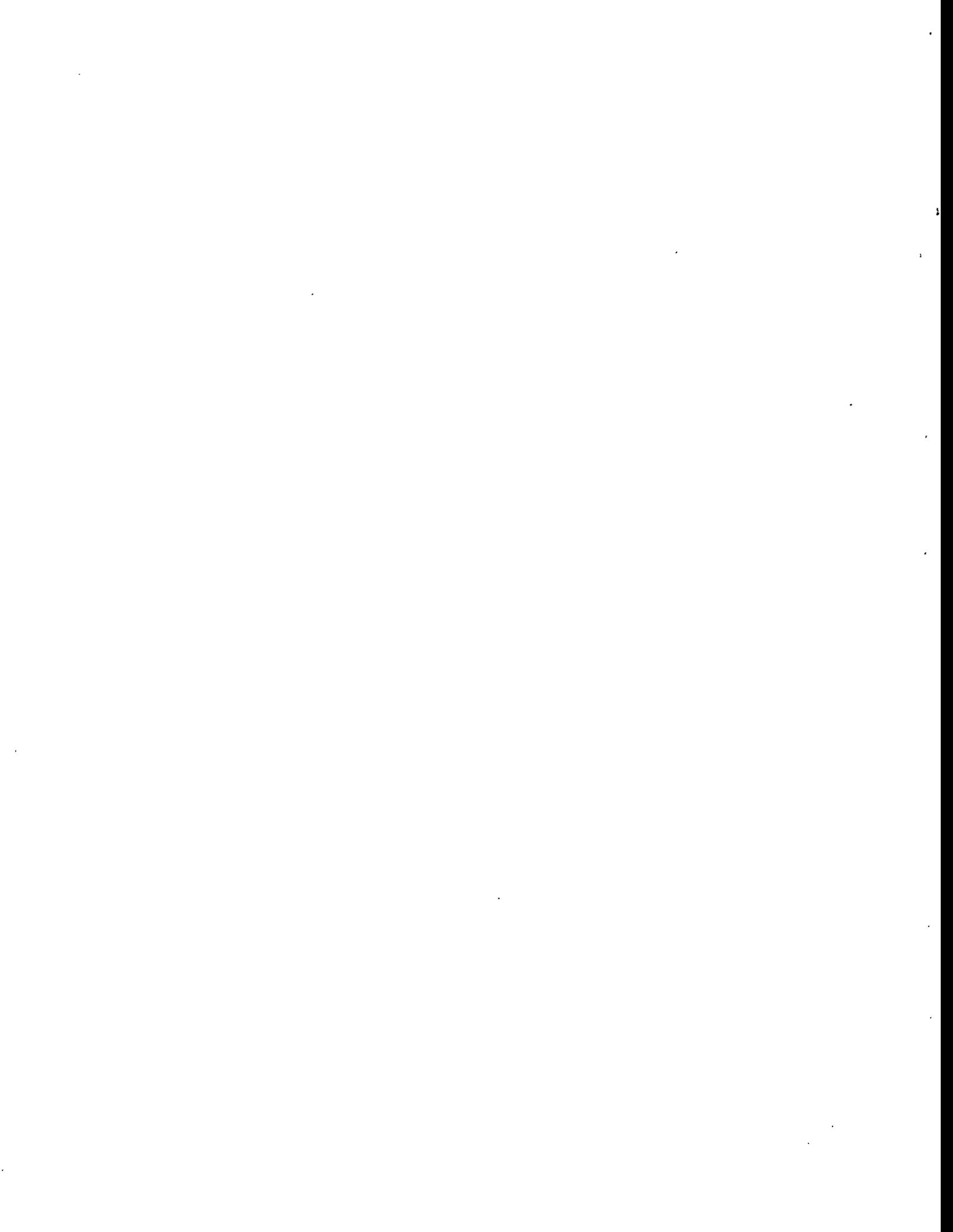
9

ADENDUM



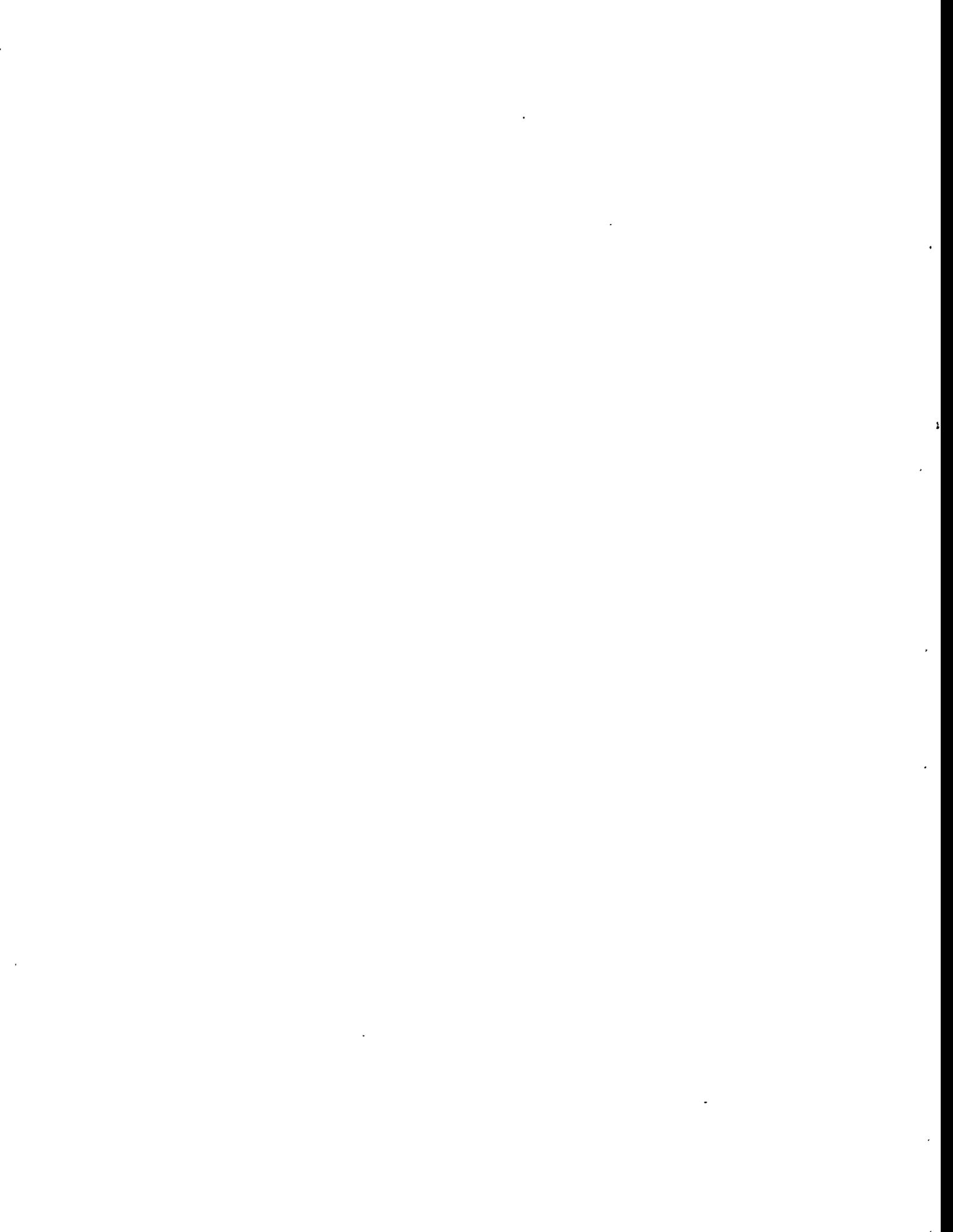
El mismo ejercicio que presentamos en este informe fue repetido bajo las mismas condiciones pero considerando que la información secundaria sólo era disponible en un 1% de las celdas del acuífero. Queremos puntualizar en este apartado algunos de los resultados a los que hemos llegado en ese ejercicio. En el caso de los métodos que requieren conocer la variable secundaria en cada nodo donde se quiere estimar la primaria, es necesario implementar un paso intermedio antes de proceder a la estimación/simulación. Se trata de krigear los datos de la variable secundaria sobre

todo el dominio de modelización. Después, al utilizar este campo interpolado para estimar la variable de interés, los campos resultantes se ven afectados por esta interpolación. Es en estos casos en donde los algoritmos no paramétricos como el krigado/cokrigado indicador muestran un mejor comportamiento que los demás. Particularmente robusto resulta el algoritmo de Markov-Bayes. En Cassiraga y Gómez-Hernández (1995) puede encontrarse un caso en donde la información blanda esta disponible en aproximadamente el 1% de los nodos de la malla a estimar.



10

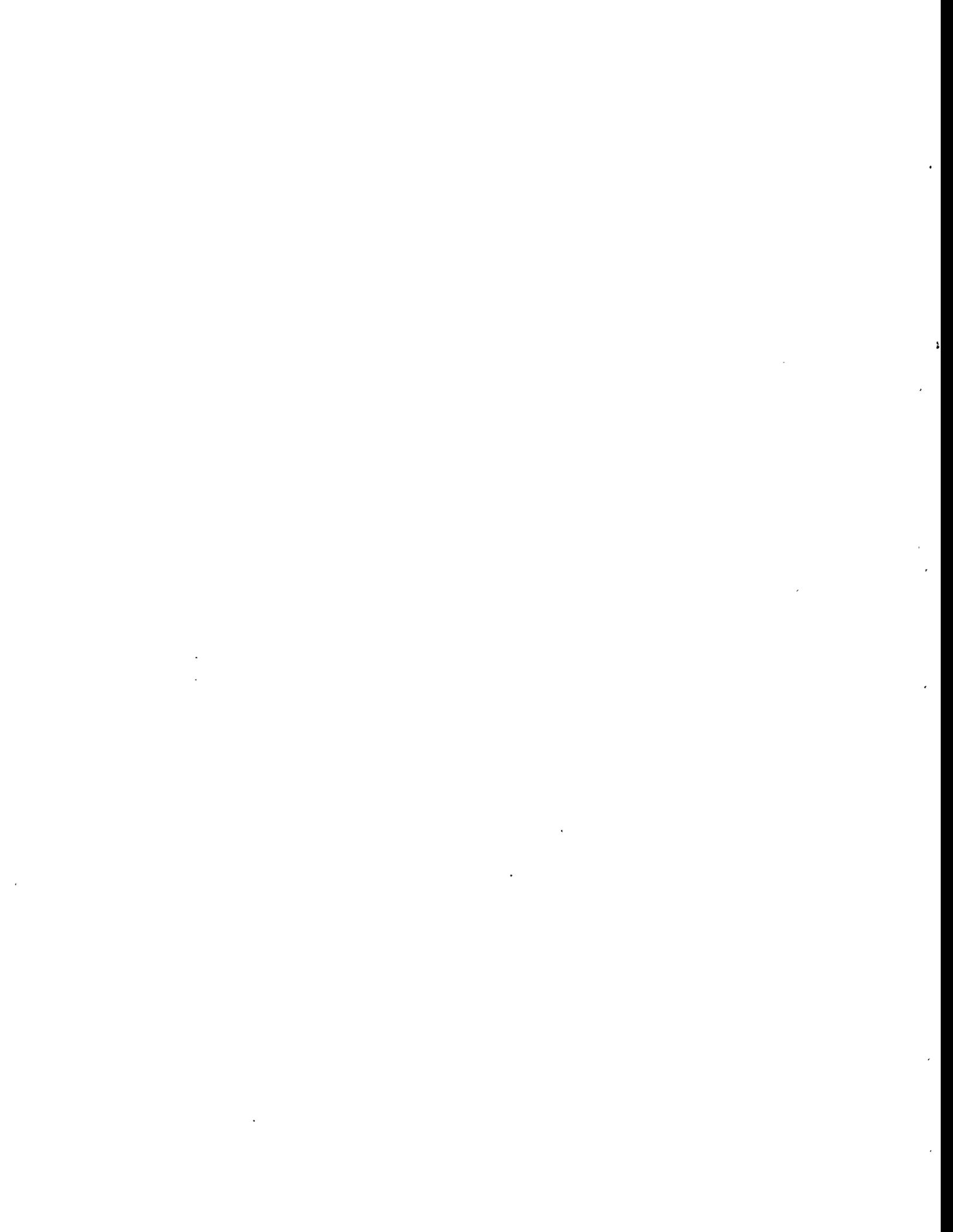
REFERENCIAS



- Aboufirassi, M. and Mariño, M. A. [1984]. Cokriging of transmissivities from field measurements of transmissivities and specific capacity. *J. Math. Geol.*, 16 (1), 19-35.
- Ahmed, S. and De Marsily, G. [1987]. Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity. *Water Resources Research*, 23 (9), 1717-1737.
- Ahmed, S., De Marsily, G. and Talbot, A. [1988]. Combined use of hydraulic and electrical properties of an aquifer in a geostatistical estimation of transmissivity. *Ground Water*, 26 (1), 78-86.
- Alabert, F. G. [1987]. *Stochastic imaging of spatial distributions using hard and soft data*. Master's thesis, Stanford University, CA, Branner Earth Sciences Library, 169 p.
- Almeida, A. S. [1993a]. Joint simulation of multiple variables with a Markov-type coregionalization model. In SCRF report 6, May 1993.
- Almeida, A. S. [1993b]. *Joint simulation of multiple variables with a Markov-type coregionalization model*. Ph. D. thesis, Stanford University, 200 p.
- Almeida, A. S., Frykman, P., Srivastava, M. and Journel, A. G. [1993]. Geostatistical modelling of the Dan field. In SCRF report 6, May 1993.
- Almeida, A. S. and Journel, A. G. [1994]. Joint simulation of multiple variables with a Markov-type coregionalization model. *Mathematical Geology*, 26 (5), 565-588.
- Araktingi, U. G., Bashore, W. M., Tran, T. T. B. and Hewett, T. A. [1992]. Integration of seismic and well log data in reservoir modeling. In SCRF report 5, May 1992.
- Asli, M. and Marcotte, D. [1995]. Comparison of approaches to spatial estimation in a bivariate context. *Mathematical Geology*, 27 (5), 641-658.
- Bardossy, A., Bogardi, I. and Kelly, W. E. [1986]. Geostatistical analysis of geoelectric estimates for specific capacity. *Journal of Hydrology*, 84, 81-95.
- Bourgault, G., Journel, A., Lesch, S. M., Rhodes, J. D. and Corwin, D. L. [1995]. Geostatistical analysis of a soil salinity data set. In Proceedings of the 1995 Bouyoucos Conference, Application of GIS to the Modeling of Non-Point Source Pollutants in the Vadose Zone, May 1-3, Riverside, CA.
- Bourgault, G. and Journel, A. [1995]. Unsmoothed estimation with dense secondary information using probability field technique. In Computer Applications in the Mineral Industry, Hani Sabri Mitri de., CAMI'95, Montreal, October 22-25, 95-102.
- Box, G. E. P. and Tiao, G. C. [1973]. *Bayesian inference in statistical analysis*. Addison-Wesley Publishing Company.
- Cassiraga, E. F. y Gómez Hernández, J. J. [1994]. Relación entre parámetros geofísicos e hidrogeológicos. Una revisión de literatura. *Publicación técnica 11/94 ENRESA*.
- Cassiraga, E. F. y Gómez Hernández, J. J. [1995]. Estimación de parámetros hidrogeológicos incorporando medidas geofísicas. En VI Simposio de Hidrogeología, Sevilla, 23-27 de octubre, tomo XIX, 85-99.
- Chu, J., Xu, W., Zhu, H. and Journel, A. G. [1991]. The Amoco case study. In SCRF report 4, May 1991.
- Chu, J., Xu, W., and Journel, A. G. [1992]. The Amoco case study-Part VI. Facing the 3rd dimension. In SCRF report 5, May 1992.
- Deutsch, C. V. [1991]. The relationship between universal kriging, kriging with an external drift, and cokriging. In SCRF report 4, May 1991.
- Deutsch, C. V. [1992]. *Annealing techniques applied to reservoir modeling and the integration of geological and engineering (well test) data*. Ph. D. thesis, Stanford University, 306 p.
- Deutsch, C. V. and Journel, A. G. [1991]. The application of simulated annealing to stochastic reservoir modeling. In SCRF report 4, May 1991.
- Deutsch, C. V. and Journel, A. G. [1992a]. *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press.
- Deutsch, C. V. and Journel, A. G. [1992b]. Annealing technics applied to the integration of geological and engineering data. In SCRF report 5, May 1992.

- Deutsch, C. V. and Cockerham, P. W. [1994]. Practical considerations in the application of simulated annealing to stochastic simulation. *Mathematical Geology*, 26 (1), 67-82.
- Doyen, P. [1988]. Porosity from seismic data: a geostatistical approach. *Geophysics*, 53 (10), 1263-1275.
- Farmer, C. [1989]. Numerical rocks. In Joint IMA/SPE European Conference on the Mathematics of Oil Recovery, Robinson College, Cambridge University.
- Froidevaux, R. [1992]. Probability field simulation. In *Geostatistics Troia'92*, proceedings of the Fourth Geostatistics Congress, A. Soares (ed.), Kluwer Academic Publishers, Dordrecht, Holland.
- Frykman, P. (ed.) [1994]. Stochastic modelling of petrophysical properties in chalk. *DGU Service report n° 35*, ENS Journal n° 1313/92-0006, Completion report to the Ministry of Energy, Ministry of the Environment, Geological Survey of Denmark.
- Gómez Hernández, J. J. [1991]. Geoestadística para el análisis de riesgos. Una introducción a la geoestadística no paramétrica. *Publicación técnica 04/91 ENRESA*.
- Gómez-Hernández, J. J. (1992). "Descripción de la base de datos Walker". *Publicación Técnica 13/92*, ENRESA.
- Gómez Hernández, J. J. [1993a]. Programas complementarios para el análisis estocástico del transporte de radionucleidos. *Publicación técnica 08/93 ENRESA*.
- Gómez Hernández, J. J. [1993b]. Métodos para la evaluación de la incertidumbre espacial. En *Primeras Jornadas de I + D en la Gestión de Residuos Radiactivos*, 21-23 abril, 1993, tomo II, publicado por ENRESA, 145-157.
- Gómez-Hernández, J. J. and Journel, A. J. [1993]. Joint sequential simulation of multiGaussian fields. In *Geostatistics Troia '92*, Part 1, A. Soares (ed.), Kluwer Academic Publ., 85-94.
- Gómez-Hernández, J. J. and Cassiraga, E. F. [1994]. Theory and practice of sequential simulation. In M. Armstrong and P. A. Dowd (eds.), *Geostatistical Simulations*, 111-124.
- Gómez-Hernández, J. J. and Wen, X.-H. [1994]. Probabilistic assessment of travel times in groundwater modeling. *Stochastic Hydrology and Hydraulics*, 8 (1), 19-55.
- Isaaks, E. and Srivastava, R. [1989]. *An introduction to applied geostatistics*. Oxford University Press, New York, 561 p.
- Istok, J. D., Smyth, J. D. and Flin, A. L. [1993]. Multivariate geostatistical analysis of groundwater contamination: a case history. *Ground Water*, 31 (1), 63-74.
- Journel, A. G. [1983]. Nonparametric estimation of spatial distributions. *Mathematical Geology*, 15 (3), 445-468.
- Journel, A. G. [1986]. Constrained interpolation and qualitative information. The soft kriging approach. *Mathematical Geology*, 18 (3), 269-286.
- Journel, A. G. [1989]. *Fundamentals of geostatistics in five lessons*. Short course in geology: volume 8, American Geophysical Union, 40 p.
- Journel, A. G. and Huijbregts, Ch. J. [1978]. *Mining Geostatistics*, Academic Press, 600 p.
- Journel, A. G. and Rossi, M. E. [1989]. When do we need a trend model?. *Mathematical Geology*, 21 (7), 715-739.
- Journel, A. G. and Zhu, H. [1990]. Integrating soft seismic data: Markov-Bayes updating, an alternative to cokriging and traditional regression. In *SCRF report 3*, May 1990.
- Journel, A. G., Xu, W. and Tran, T. [1992]. Integrating seismic data in reservoir modeling: the collocated cokriging alternative. In *SCRF report 5*, May 1992.
- Langlais, V. and Doyle, J. [1993]. Comparison of several methods of facies simulation on the Fluvial Gypsy Sandstone of Oklahoma. In *Geostatistics Troia '92*, Part 1, A. Soares (ed.), Kluwer Academic Publ., 299-310.
- Marechal, A. [1984]. Kriging seismic data in presence of faults. In G. Verly et al., editors, *Geostatistics for natural resources characterization*, 271-294, Reidel, Dordrecht, Holland.
- Mavko, G., Lucet, N. and VanSchaak, M. [1991]. Rock properties interpretation of a cross-well seismic image. In *SCRF report 4*, May 1991.
- Muñoz-Pardo, J. F. and García, R. [1989]. Estimation of the transmissivity of the Santiago

- aquifer, Chile, using different geostatistical methods. In *Groundwater Management: Quantity and Quality* (Proceedings of the Benidorm Symposium, October 1989). IAHS publ. n° 188, 77-84.
- Myers, D. E. [1982]. Matrix formulation of co-kriging. *J. Math. Geol.*, 14 (3), 249-257.
- Myers, D. E. [1984]. Cokriging-New developments. *Geostatistics for Natural Resources Characterization*, part 1, edited by G. Verly et al., NATO ASI Series, D. Reidel Publishing Company, 295-305.
- Myers, D. E. [1985]. Cokriging: Methods and alternatives. *The Role of Data in Scientific Progress*, edited by P. S. Glaeser, Elsevier Science, New York, 425-428.
- Pesti, G., Bogardi, I., Kelly, W. E. and Kalinski, R. J. [1993]. Cokriging of geoelectric and well data to define aquifer protective layers. *Ground Water*, 31 (6), 905-912.
- Srivastava, R. M. [1992]. Reservoir characterization with probability field simulation. *SPE paper 24753* presented at the 1992 Annual Technical Conference of the Society of Petroleum engineers, SPE, Richardson, Texas.
- Srivastava, R. M. [1994]. The visualization of spatial uncertainty. In *SCRF report 7*, May 1994 (only the abstract).
- Verly, G. [1993]. Sequential Gaussian cosimulation: a simulation method integrating several types of information. In *Geostatistics Troia' 92*, Part 1, A. Soares (ed.), Kluwer Academic Publ., 543-554.
- Wen, X.-H. and Kung, Ch.-S. [1993]. Stochastic simulation of solute transport in heterogeneous formations : A comparison of parametric and nonparametric geostatistical approaches. *Ground Water*, 31 (6), 953-965.
- Zhu, H. [1991]. *Modeling mixture of spatial distributions with integration of soft data*. Ph. D. thesis, Stanford University, p 250.
- Zhu, H. and Journel, A. [1992]. Formatting and integrating soft data: stochastic imaging via the Markov-Bayes algorithm. In *Geostatistics Troia' 92*, Part 1, A. Soares (ed.), Kluwer Academic Publ., 1-12.



PUBLICACIONES TECNICAS

1991

-
- 01 REVISION SOBRE LOS MODELOS NUMERICOS RELACIONADOS CON EL ALMACENAMIENTO DE RESIDUOS RADIATIVOS.
 - 02 REVISION SOBRE LOS MODELOS NUMERICOS RELACIONADO CON EL ALMACENAMIENTO DE RESIDUOS RADIATIVOS. ANEXO 1. Guía de códigos aplicables.
 - 03 PRELIMINARY SOLUBILITY STUDIES OF URANIUM DIOXIDE UNDER THE CONDITIONS EXPECTED IN A SALINE REPOSITORY.
 - 04 GEOESTADISTICA PARA EL ANALISIS DE RIESGOS. Una introducción a la Geoestadística no paramétrica.
 - 05 SITUACIONES SINOPTICAS Y CAMPOS DE VIENTOS ASOCIADOS EN "EL CABRIL".
 - 06 PARAMETERS, METHODOLOGIES AND PRIORITIES OF SITE SELECTION FOR RADIOACTIVE WASTE DISPOSAL IN ROCK SALT FORMATIONS.

1992

-
- 01 STATE OF THE ART REPORT: DISPOSAL OF RADIOACTIVE WASTE IN DEEP ARGILLACEOUS FORMATIONS.
 - 02 ESTUDIO DE LA INFILTRACION A TRAVES DE LA COBERTERA DE LA FUA.
 - 03 SPANISH PARTICIPATION IN THE INTERNATIONAL INTRAVAL PROJECT.
 - 04 CARACTERIZACION DE ESMECTITAS MAGNESICAS DE LA CUENCA DE MADRID COMO MATERIALES DE SELLADO. Ensayos de alteración hidrotermal.
 - 05 SOLUBILITY STUDIES OF URANIUM DIOXIDE UNDER THE CONDITIONS EXPECTED IN A SALINE REPOSITORY. Phase II
 - 06 REVISION DE METODOS GEOFISICOS APPLICABLES AL ESTUDIO Y CARACTERIZACION DE EMPLAZAMIENTOS PARA ALMACENAMIENTO DE RESIDUOS RADIATIVOS DE ALTA ACTIVIDAD EN GRANITOS, SALES Y ARCILLAS.
 - 07 COEFICIENTES DE DISTRIBUCION ENTRE RADIONUCLEIDOS.
 - 08 CONTRIBUTION BY CTN-UPM TO THE PSACOIN LEVEL-S EXERCISE.
 - 09 DESARROLLO DE UN MODELO DE RESUSPENSION DE SUELOS CONTAMINADOS. APLICACION AL AREA DE PALOMARES.
 - 10 ESTUDIO DEL CODIGO FFSM PARA CAMPO LEJANO. IMPLANTACION EN VAX.
 - 11 LA EVALUACION DE LA SEGURIDAD DE LOS SISTEMAS DE ALMACENAMIENTO DE RESIDUOS RADIATIVOS. UTILIZACION DE METODOS PROBABILISTAS.
 - 12 METODOLOGIA CANADIENSE DE EVALUACION DE LA SEGURIDAD DE LOS ALMACENAMIENTOS DE RESIDUOS RADIATIVOS.
 - 13 DESCRIPCION DE LA BASE DE DATOS WALKER.

1993

-
- 01 INVESTIGACION DE BENTONITAS COMO MATERIALES DE SELLADO PARA ALMACENAMIENTO DE RESIDUOS RADIATIVOS DE ALTA ACTIVIDAD. ZONA DE CABO DE GATA, ALMERIA.

- 02 TEMPERATURA DISTRIBUTION IN A HYPOTHETICAL SPENT NUCLEAR FUEL REPOSITORY IN A SALT DOME.
- 03 ANALISIS DEL CONTENIDO EN AGUA EN FORMACIONES SALINAS. Su aplicación al almacenamiento de residuos radiactivos
- 04 SPANISH PARTICIPATION IN THE HAW PROJECT. Laboratory Investigations on Gamma Irradiation Effects in Rock Salt.
- 05 CARACTERIZACION Y VALIDACION INDUSTRIAL DE MATERIALES ARCILLOSOS COMO BARRERA DE INGENIERIA.
- 06 CHEMISTRY OF URANIUM IN BRINES RELATED TO THE SPENT FUEL DISPOSAL IN A SALT REPOSITORY (I).
- 07 SIMULACION TERMICA DEL ALMACENAMIENTO EN GALERIA-TSS.
- 08 PROGRAMAS COMPLEMENTARIOS PARA EL ANALISIS ESTOCASTICO DEL TRANSPORTE DE RADIONUCLEIDOS.
- 09 PROGRAMAS PARA EL CALCULO DE PERMEABILIDADES DE BLOQUE.
- 10 METHODS AND RESULTS OF THE INVESTIGATION OF THE THERMOMECHANICAL BEHAVIOUR OF ROCK SALT WITH REGARD TO THE FINAL DISPOSAL OF HIGH-LEVEL RADIOACTIVE WASTES.

1994

-
- 01 MODELO CONCEPTUAL DE FUNCIONAMIENTO DE LOS ECOSISTEMAS EN EL ENTORNO DE LA FABRICA DE URANIO DE ANDUJAR.
 - 02 CORROSION OF CANDIDATE MATERIALS FOR CANISTER APPLICATIONS IN ROCK SALT FORMATIONS.
 - 03 STOCHASTIC MODELING OF GROUNDWATER TRAVEL TIMES
 - 04 THE DISPOSAL OF HIGH LEVEL RADIOACTIVE WASTE IN ARGILLACEOUS HOST ROCKS. Identification of parameters, constraints and geological assessment priorities.
 - 05 EL OESTE DE EUROPA Y LA PENINSULA IBERICA DESDE HACE -120.000 AÑOS HASTA EL PRESENTE. Isostasia glacial, paleogeografías y paleotemperaturas.
 - 06 ECOLOGIA EN LOS SISTEMAS ACUATICOS EN EL ENTORNO DE EL CABRIL.
 - 07 ALMACENAMIENTO GEOLOGICO PROFUNDO DE RESIDUOS RADIATIVOS DE ALTA ACTIVIDAD (AGP). Conceptos preliminares de referencia.
 - 08 UNIDADES MOVILES PARA CARACTERIZACION HIDROGEOQUIMICA
 - 09 EXPERIENCIAS PRELIMINARES DE MIGRACION DE RADIONUCLEIDOS CON MATERIALES GRANITICOS. EL BERROCAL, ESPAÑA.
 - 10 ESTUDIOS DE DESEQUILIBRIOS ISOTOPICOS DE SERIES RADIATIVAS NATURALES EN UN AMBIENTE GRANITICO: PLUTON DE EL BERROCAL (TOLEDO).
 - 11 RELACION ENTRE PARAMETROS GEOFISICOS E HIDROGEOLOGICOS. Una revisión de literatura.
 - 12 DISEÑO Y CONSTRUCCION DE LA COBERTURA MULTICAPA DEL DIQUE DE ESTERILES DE LA FABRICA DE URANIO DE ANDUJAR.

1995

-
- 01 DETERMINACION DEL MODULO DE ELASTICIDAD DE FORMACIONES ARCILLOSAS PROFUNDAS.

- 02 *UO₂ LEACHING AND RADIONUCLIDE RELEASE MODELLING UNDER HIGH AND LOW IONIC STRENGTH SOLUTION AND OXIDATION CONDITIONS.*
- 03 *THERMO-HYDRO-MECHANICAL CHARACTERIZATION OF THE SPANISH REFERENCE CLAY MATERIAL FOR ENGINEERED BARRIER FOR GRANITE AND CLAY HLW REPOSITORY: LABORATORY AND SMALL MOCK UP TESTING.*
- 04 *DOCUMENTO DE SINTESIS DE LA ASISTENCIA GEOTECNICA AL DISEÑO AGP-ARCILLA. Concepto de referencia.*
- 05 *DETERMINACION DE LA ENERGIA ACUMULADA EN LAS ROCAS SALINAS FUERTEMENTE IRRADIADAS MEDIANTE TECNICAS DE TERMOLUMINISCENCIA. Aplicación al análisis de repositorios de residuos radiactivos de alta actividad.*
- 06 *PREDICCIÓN DE FENOMENOS DE TRANSPORTE EN CAMPO PRÓXIMO Y LEJANO. Interacción en fases sólidas.*
- 07 *ASPECTOS RELACIONADOS CON LA PROTECCION RADIOLOGICA DURANTE EL DESMANTELAMIENTO Y CLAUSURA DE LA FABRICA DE ANDUJAR.*
- 08 *ANALYSIS OF GAS GENERATION MECHANISMS IN UNDERGROUND RADIOACTIVE WASTE REPOSITORIES. (Pegase Project).*
- 09 *ENSAYOS DE LIXIVIACION DE EMISORES BETA PUROS DE LARGA VIDA.*
- 10 *2º PLAN DE I+D. DESARROLLOS METODOLOGICOS, TECNOLOGICOS, INSTRUMENTALES Y NUMERICOS EN LA GESTION DE RESIDUOS RADIATIVOS.*
- 11 *PROYECTO AGP- ALMACENAMIENTO GEOLOGICO PROFUNDO. FASE 2.*
- 12 *IN SITU INVESTIGATION OF THE LONG-TERM SEALING SYSTEM AS COMPONENT OF DAM CONSTRUCTION (DAM PROJECT). Numerical simulator: Code-Bright.*

1996

- 01 *DESARROLLO DE UN PROGRAMA INFORMATICO PARA EL ASESORAMIENTO DE LA OPERACION DE FOCOS EMISORES DE CONTAMINANTES GASEOSOS.*
- 02 *FINAL REPORT OF PHYSICAL TEST PROGRAM CONCERNING SPANISH CLAYS (SAPONITES AND BENTONITES).*
- 03 *APORTACIONES AL CONOCIMIENTO DE LA EVOLUCION PALEOCLIMATICA Y PALEOAMBIENTAL EN LA PENINSULA IBERICA DURANTE LOS DOS ULTIMOS MILLONES DE AÑOS A PARTIR DEL ESTUDIO DE TRAVERTINOS Y ESPELEOTEMAS.*

PUBLICACIONES NO PERIODICAS

1992

PONENCIAS E INFORMES, 1988-1991.
SEGUNDO PLAN DE I+D, 1991-1995. TOMOS I, II Y III.
SECOND RESEARCH AND DEVELOPMENT PLAN, 1991-1995, VOLUME I.

1993

SEGUNDO PLAN DE I+D. INFORME ANUAL 1992.
PRIMERAS JORNADAS DE I+D EN LA GESTION DE RESIDUOS RADIATIVOS. TOMOS I Y II.

1994

SEGUNDO PLAN I+D 1991-1995. INFORME ANUAL 1993.

1995

TERCER PLAN DE I+D 1995-1999.
SEGUNDAS JORNADAS DE I+D. EN LA GESTION DE RESIDUOS RADIATIVOS. TOMOS I Y II.

Edita:

enresa

empresa nacional de residuos radiactivos, s.a.

Diseño y coordinación editorial: TransEdit
Imprime: L & J DUQUE, S.L.
ISSN: 1134-380X
D.L.: M-22623-1996
Junio de 1996